(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 December 2002 (05.12.2002)

PCT

(10) International Publication Number
WO 02/096951 A1

(51) International Patent Classification[7]: C07K 17/00, G01N 33/53, A61K 39/00

(21) International Application Number: PCT/US02/16676

(22) International Filing Date: 24 May 2002 (24.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/293,723	25 May 2001 (25.05.2001)	US
60/295,257	1 June 2001 (01.06.2001)	US
60/297,220	8 June 2001 (08.06.2001)	US
60/300,526	21 June 2001 (21.06.2001)	US
60/301,874	29 June 2001 (29.06.2001)	US
60/359,413	22 February 2002 (22.02.2002)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **TANG, Tom, Y.** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **DUGGAN, Brendan, M.** [AU/US]; 243 Buena Vista Avenue # 306, Sunnyvale, CA 94086 (US). **WARREN, Bridget, A.** [US/US]; 2250 Homestead Court # 2, Los Altos, CA 94024 (US). **BANDMAN, Olga** [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). **RICHARDSON, Thomas, W.** [US/US]; 616 Canyon Road #107, Redwood City, CA 94062 (US). **BURFORD, Neil** [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). **SANJANWALA, Bharati** [IN/US]; 210 Silvia Court, Los Altos, CA 94024 (US). **BECHA, Shanya, D.** [US/US]; 21062 Gary Drive # 117, Castro Valley, CA 94546 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). **YANG, Junming** [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US). **TRAN, Uyen, K.** [US/US]; 2638 Mabury Square, San Jose, CA 95133 (US). **HAFALIA, April, J.A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **GRIFFIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **SWARNAKAR, Anita** [CA/US]; 8 Locksley Avenue #5D, San Francisco, CA 94122 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **RECIPON, Shirley, A.** [US/US]; 85 Fortuna Avenue, San Francisco, CA 94115 (US). **KHAN, Farrah, A.** [IN/US]; 9445 Harrison Street, Des Plaines, IL 60016 (US). **LEE, Ernestine, A.** [US/US]; 20523 Crow Creek Road, Castro Valley, CA 94552 (US). **YUE, Huibin** [US/US]; 1170 South Stelling Road, Cupertino, CA 95014 (US). **LU, Dyung Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street A # 205, San Leandro, CA 94577 (US). **THANGAVELU, Kavitha** [IN/US]; 1950 Montecito Avenue #23, Mountain View, CA 94043 (US). **ARVIZU, Chandra, S.** [US/US]; 1706 Morocco Drive, San Jose, CA 95125 (US). **XU, Yuming** [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). **ISON, Craig, H.** [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US). **HUANG, Jiaqi** [CN/US]; 210 Calderon Avenue #122, Mountain View, CA 94041 (US). **DING, Li** [CN/US]; 3353 Alma Street # 146, Palo Alto, CA 94306 (US). **HONCHELL, Cynthia, D.** [US/US]; 400 Laurel Street #203, San Carlos, CA 94070 (US). **BOROWSKY, Mark, L.** [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). **EMERLING, Brooke, M.** [US/US]; 1735 Woodland Avenue # 71, Palo Alto, CA 94303 (US). **PETERSON, David, P.** [US/US]; 970 Cherry Avenue, San Jose, CA 95126 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **MASON, Patricia, M.** [US/US]; 360 Clarke Lane, Morgan Hill, CA 95037 (US). **ZEBARJADIAN, Yeganeh** [IR/US]; 830 Junipero Serra Boulevard, San Francisco, CA 94127 (US). **AZIMZAI, Yalda** [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). **STUVE, Laura, L.** [US/US]; 14630 Stetson Road, Los Gatos, CA 95030 (US). **KAMIGAKI, Laura, L.** [US/US]; 6921 Corte Mateo, Pleasanton, CA 94566 (US). **BARROSO, Ines** [PT/GB]; 38 Eden Street, Cambridge, Antrim CB1 1EL (GB). **LEE, Sally** [US/US]; 3643 26th Street, San Francisco, CA 94110 (US). **KABLE, Amy, E.** [US/US]; 2345 Polk Street #4, San Francisco, CA 94109 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,

[Continued on next page]

(54) Title: MOLECULES FOR DISEASE DETECTION AND TREATMENT

(57) Abstract: The invention provides human molecules for disease detection and treatment (MDDT)and polynucleotides which identify and encode MDDT. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methodsfor diagnosing, treating, or preventing disorders associated with aberrant expression of MDDT.

WO 02/096951 A1

CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

MOLECULES FOR DISEASE DETECTION AND TREATMENT

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of molecules for disease detection and treatment and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative, autoimmune/inflammatory, developmental, and neurological disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of molecules for disease detection and treatment.

BACKGROUND OF THE INVENTION

The human genome is comprised of thousands of genes, many encoding gene products that function in the maintenance and growth of the various cells and tissues in the body. Aberrant expression or mutations in these genes and their products is the cause of, or is associated with, a variety of human diseases such as cancer and other cell proliferative disorders. The identification of these genes and their products is the basis of an ever-expanding effort to find markers for early detection of diseases, and targets for their prevention and treatment.

It is estimated that only 2% of mammalian DNA encodes proteins, and only a small fraction of the genes that encode proteins are actually expressed in a particular cell at any time. The various types of cells in a multicellular organism differ dramatically both in structure and function, and the identity of a particular cell is conferred by its unique pattern of gene expression. In addition, different cell types express overlapping but distinctive sets of genes throughout development. Cell growth and proliferation, cell differentiation, the immune response, apoptosis, and other processes that contribute to organismal development and survival are governed by regulation of gene expression. Appropriate gene regulation also ensures that cells function efficiently by expressing only those genes whose functions are required at a given time. Factors that influence gene expression include extracellular signals that mediate cell-cell communication and coordinate the activities of different cell types. Gene expression is regulated at the level of DNA and RNA transcription, and at the level of mRNA translation.

Cancer represents a type of cell proliferative disorder that affects nearly every tissue in the body. A wide variety of molecules, either aberrantly expressed or mutated, can be the cause of, or involved with, various cancers because tissue growth involves complex and ordered patterns of cell proliferation, cell differentiation, and apoptosis. Cell proliferation must be regulated to maintain both the number of cells and their spatial organization. This regulation depends upon the appropriate

expression of proteins which control cell cycle progression in response to extracellular signals such as growth factors and other mitogens, and intracellular cues such as DNA damage or nutrient starvation. Molecules which directly or indirectly modulate cell cycle progression fall into several categories, including growth factors and their receptors, second messenger and signal transduction proteins, oncogene products, tumor-suppressor proteins, and mitosis-promoting factors.

Aberrant expression or mutations in genes and their products may cause, or increase susceptibility to, a variety of human diseases such as cancer and other cell proliferative disorders. The identification of these genes and their products is the basis of an ever-expanding effort to find markers for early detection of diseases and targets for their prevention and treatment. For example, cancer represents a type of cell proliferative disorder that affects nearly every tissue in the body. The development of cancer, or oncogenesis, is often correlated with the conversion of a normal gene into a cancer-causing gene, or oncogene, through abnormal expression or mutation. Oncoproteins, the products of oncogenes, include a variety of molecules that influence cell proliferation, such as growth factors, growth factor receptors, intracellular signal transducers, nuclear transcription factors, and cell-cycle control proteins. In contrast, tumor-suppressor genes are involved in inhibiting cell proliferation. Mutations which reduce or abrogate the function of tumor-suppressor genes result in aberrant cell proliferation and cancer. Thus a wide variety of genes and their products have been found that are associated with cell proliferative disorders such as cancer, but many more may exist that are yet to be discovered.

Mammalian peripheral blood comprises cells of the erythroid, myeloid, and lymphoid lineages. Each lineage is derived from a pluripotent stem cell which, upon exposure to various molecules and other types of cells, differentiate into effector cells which migrate into the blood and other organs. These cells include red blood cells and platelets (erythroid), macrophages and granulocytes (myeloid), and T and B lymphocytes (lymphoid). Myeloid and lymphoid cells mediate immune responses to pathogens such as bacteria, parasites, and viruses.

Functional interaction of the cell types involved in immune responses involves transfer of signals via soluble messenger molecules known as cytokines. Both hematopoietic cells and non-hematopoietic cells produce cytokines which stimulate the activation, differentiation and proliferation of T cells, B cells, macrophages, and granulocytes during an active immune response. Cytokines bind to specific receptors expressed on cellular membranes and transduce a signal through the cell. Depending on the type of cytokine and the cell to which it binds, this signal initiates activation, differentiation, growth, and/or apoptosis.

T cells, which respond to and produce a variety of cytokines, are divided into two major groups, $CD4^+$ T helper (Th) cells, and $CD8^+$ cytotoxic T lymphocytes (CTL). Immune responses are

primarily regulated by $CD4^+$ Th cells which fall into two subclasses based on the kinds of cytokines they secrete. Th1 cells secrete primarily IL-2 and IFN-γ; regulate the responses of CTLs, B cells, and macrophages; and orchestrate the removal of intracellular pathogens. In contrast, Th2 cells secrete primarily IL-4 and IL-10 and promote the development of certain antibody responses such as IgG1, IgA, and IgE. In addition, Th2 cells remove extracellular pathogens, which include various bacteria and parasites. (See, e.g., Morel and Oriss (1998) Crit. Rev. Immunol. 18:275-303.) Studies have shown that the Th1 cytokine response predominates in organ-specific autoimmune disorders such as insulin-dependent diabetes mellitus (IDDM), multiple sclerosis (MS), rheumatoid arthritis (RA), and Crohn's disease. A Th1 response also predominates in acute allograft rejection, eradication of tumors, and unexplained recurrent abortions. Th2 responses predominate in allergy and other atopic disorders, transplantation tolerance, chronic graft versus host disease (GVHD), and systemic autoimmune disease such as systemic lupus erythmatosus (Romangnani et al. (1997) Int. Arch. Allergy Immunol. 113:153-156). Genes affected by these molecules may reasonably be expected to be markers of immune cell development, function, and activity.

Tumor necrosis factor (TNF) α is a pleiotropic cytokine that mediates immune regulation and inflammatory responses. TNF-α-related cytokines generate partially overlapping cellular responses, including differentiation, proliferation, nuclear factor-κB (NF-κB) activation, and cell death, by triggering the aggregation of receptor monomers (Smith, C.A. et al. (1994) Cell 76:959-962). The cellular responses triggered by TNF-α are initiated through its interaction with distinct cell surface receptors (TNFRs). Treatment of confluent cultures of vascular smooth muscle cells (SMCs) with TNF-α suppresses the incorporation of [^{3}H]proline into both collagenase-digestible proteins (CDP) and noncollagenous proteins (NCP). Such suppression by TNF-α is not observed in confluent bovine aortic endothelial cells and human fibroblastic IMR-90 cells. TNF-α decreases the relative proportion of collagen types IV and V suggesting that TNF-α modulates collagen synthesis by SMCs depending on their cell density and therefore may modify formation of atherosclerotic lesions (Hiraga, S. et al. (2000) Life Sci. 66:235-244). Primary human endothelial cell lines such as human umbilical vein endothelial cells (HUVECs) have been used as an experimental model for investigating in vitro the role of the endothelium in human vascular biology. Activation of the vascular endothelium is considered to be a central event in a wide range of both physiological and pathophysiological processes, such as vascular tone regulation, coagulation and thrombosis, atherosclerosis, and inflammation.

DNA-based arrays can provide an efficient, high-throughput method to examine gene expression and genetic variability. For example, SNPs, or single nucleotide polymorphisms, are the most common type of human genetic variation. DNA-based arrays can dramatically accelerate the

discovery of SNPs in hundreds and even thousands of genes. Likewise, such arrays can be used for SNP genotyping in which DNA samples from individuals or populations are assayed for the presence of selected SNPs. These approaches will ultimately lead to the systematic identification of all genetic variations in the human genome and the correlation of certain genetic variations with disease susceptibility, responsiveness to drug treatments, and other medically relevant information. (See, for example, Wang, D.G. et al. (1998) Science 280:1077-1082.)

DNA-based array technology is especially important for the rapid analysis of global gene expression patterns. For example, genetic predisposition, disease, or therapeutic treatment may directly or indirectly affect the expression of a large number of genes in a given tissue. In this case, it is useful to develop a profile, or transcript image, of all the genes that are expressed and the levels at which they are expressed in that particular tissue. A profile generated from an individual or population affected with a certain disease or undergoing a particular therapy may be compared with a profile generated from a control individual or population. Such analysis does not require knowledge of gene function, as the expression profiles can be subjected to mathematical analyses which simply treat each gene as a marker. Furthermore, gene expression profiles may help dissect biological pathways by identifying all the genes expressed, for example, at a certain developmental stage, in a particular tissue, or in response to disease or treatment. (See, for example, Lander, E.S. et al. (1996) Science 274:536-539.)

Dendritic cells (DC) are antigen presenting cells (APC) that play a key role in the primary immune response because of their unique ability to present antigens to naive T-cells. In addition, DC differentiate into separate subsets of mature immune cells that sustain and regulate immune responses following initial contact with antigen. DC subsets include those that preferentially induce particular T helper 1 (Th1) or T helper 2 (Th2) responses and those that regulate B cell responses. Moreover, DC are being used with increasing frequency to manipulate immune responses, either to downregulate aberrant autoimmune response or to enhance vaccination or tumor-specific response.

DC are functionally specialized in correlation with their particular differentiation state. CD34+ myeloid cells found in the bone marrow mature in response to signals into CD14+ CD11c+ monocytes. An innate or antigen non-specific response takes place initially when monocytes circulate to nonlymphoid tissues and respond to lipopolysaccharide (LPS), a bacterially-derived mitogen, and viruses. Such direct encounters with antigen cause secretion of pro-inflammatory cytokines that attract and regulate natural killer cells, macrophages, and eosinophils in the first line of defense against invading pathogens. Monocytes then mature into DC, which efficiently capture antigen through endocytosis and antigen-receptor uptake. Antigen processing and presentation trigger activation and differentiation into mature DC that express MHC class II molecules on the cell surface and efficiently

activate T-cells, initiating antigen-specific T-cell and B-cell responses. In turn, T-cells activate DC through CD40 ligand - CD40 interactions, which stimulate expression of the costimulatory molecules CD80 and CD86, the latter most potent in amplifying T-cell responses. DC interaction via CD40 with T cells also stimulates the production of inflammatory cytokines such as TNF alpha and IL-1. Engagement of RANK, a member of the TNF receptor family by its ligand, TRANCE, which is expressed on activated T cells, enhances the survival of DC through inhibition of apoptosis, thereby enhancing T cell activation. The maturation and differentiation of monocytes into mature DC links the antigen non-specific innate immune response to the antigen-specific adaptive immune response.

Certain genes are known to be associated with diseases because of their chromosomal location, such as the genes in the myotonic dystrophy (DM) regions of mouse and human. The mutation underlying DM has been localized to a gene encoding the DM-kinase protein, but another active gene, DMR-N9, is in close proximity to the DM-kinase gene (Jansen, G. et al. (1992) Nat. Genet. 1:261-266). DMR-N9 encodes a 650 amino acid protein that contains WD repeats, motifs found in cell signaling proteins. DMR-N9 is expressed in all neural tissues and in the testis, suggesting a role for DMR-N9 in the manifestation of mental and testicular symptoms in severe cases of DM (Jansen, G. et al. (1995) Hum. Mol. Genet. 4:843-852).

Other genes are identified based upon their expression patterns or association with disease syndromes. For example, autoantibodies to subcellular organelles are found in patients with systemic rheumatic diseases. A recently identified protein, golgin-67, belongs to a family of Golgi autoantigens having alpha-helical coiled-coil domains (Eystathioy, T. et al. (2000) J. Autoimmun. 14:179-187). The Stac gene was identified as a brain specific, developmentally regulated gene. The Stac protein contains an SH3 domain, and is thought to be involved in neuron-specific signal transduction (Suzuki, H. et al. (1996) Biochem. Biophys. Res. Commun. 229:902-909).

Intracellular Signaling

Cell-cell communication is essential for the growth, development, and survival of multicellular organisms. Cells communicate by sending and receiving molecular signals. An example of a molecular signal is a growth factor, which binds and activates a specific transmembrane receptor on the surface of a target cell. The activated receptor transduces the signal intracellularly, thus initiating a cascade of biochemical reactions that ultimately affect gene transcription and cell cycle progression in the target cell.

Intracellular signaling is the process by which cells respond to extracellular signals (hormones, neurotransmitters, growth and differentiation factors, etc.) through a cascade of biochemical reactions that begins with the binding of a signaling molecule to a cell membrane receptor and ends with the activation of an intracellular target molecule. Intermediate steps in the process involve the activation

of various cytoplasmic proteins by phosphorylation via protein kinases, and their deactivation by protein phosphatases, and the eventual translocation of some of these activated proteins to the cell nucleus where the transcription of specific genes is triggered. The intracellular signaling process regulates all types of cell functions including cell proliferation, cell differentiation, and gene transcription, and involves a diversity of molecules including protein kinases and phosphatases, and second messenger molecules such as cyclic nucleotides, calcium-calmodulin, inositol, and various mitogens that regulate protein phosphorylation.

Cells also respond to changing conditions by switching off signals. Many signal transduction proteins are short-lived and rapidly targeted for degradation by covalent ligation to ubiquitin, a highly conserved small protein. Cells also maintain mechanisms to monitor changes in the concentration of denatured or unfolded proteins in membrane-bound extracytoplasmic compartments, including a transmembrane receptor that monitors the concentration of available chaperone molecules in the endoplasmic reticulum and transmits a signal to the cytosol to activate the transcription of nuclear genes encoding chaperones in the endoplasmic reticulum.

Certain proteins in intracellular signaling pathways serve to link or cluster other proteins involved in the signaling cascade. These proteins are referred to as scaffold, anchoring, or adaptor proteins. (For review, see Pawson, T. and J.D. Scott (1997) Science 278:2075-2080.) As many intracellular signaling proteins such as protein kinases and phosphatases have relatively broad substrate specificities, the adaptors help to organize the component signaling proteins into specific biochemical pathways. Many of the above signaling molecules are characterized by the presence of particular domains that promote protein-protein interactions. A sampling of these domains is discussed below, along with other important intracellular messengers.

Intracellular Signaling Second Messenger Molecules

Protein Phosphorylation

Protein kinases and phosphatases play a key role in the intracellular signaling process by controlling the phosphorylation and activation of various signaling proteins. The high energy phosphate for this reaction is generally transferred from the adenosine triphosphate molecule (ATP) to a particular protein by a protein kinase and removed from that protein by a protein phosphatase. Protein kinases are roughly divided into two groups: those that phosphorylate serine or threonine residues (serine/threonine kinases, STK) and those that phosphorylate tyrosine residues (protein tyrosine kinases, PTK). A few protein kinases have dual specificity for serine/threonine and tyrosine residues. Almost all kinases contain a conserved 250-300 amino acid catalytic domain containing specific residues and sequence motifs characteristic of the kinase family (Hardie, G. and S. Hanks (1995) The

Protein Kinase Facts Books, Vol I:7-20, Academic Press, San Diego, CA).

STKs include the second messenger dependent protein kinases such as the cyclic-AMP dependent protein kinases (PKA), involved in mediating hormone-induced cellular responses; calcium-calmodulin (CaM) dependent protein kinases, involved in regulation of smooth muscle contraction, glycogen breakdown, and neurotransmission; and the mitogen-activated protein kinases (MAP kinases) which mediate signal transduction from the cell surface to the nucleus via phosphorylation cascades. Altered PKA expression is implicated in a variety of disorders and diseases including cancer, thyroid disorders, diabetes, atherosclerosis, and cardiovascular disease (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, New York, NY, pp. 416-431, 1887).

PTKs are divided into transmembrane, receptor PTKs and nontransmembrane, non-receptor PTKs. Transmembrane PTKs are receptors for most growth factors. Non-receptor PTKs lack transmembrane regions and, instead, form complexes with the intracellular regions of cell surface receptors. Receptors that function through non-receptor PTKs include those for cytokines and hormones (growth hormone and prolactin) and antigen-specific receptors on T and B lymphocytes. Many of these PTKs were first identified as the products of mutant oncogenes in cancer cells in which their activation was no longer subject to normal cellular controls. In fact, about one third of the known oncogenes encode PTKs, and it is well known that cellular transformation (oncogenesis) is often accompanied by increased tyrosine phosphorylation activity (Charbonneau H. and N.K. Tonks (1992) Annu. Rev. Cell Biol. 8:463-493).

An additional family of protein kinases previously thought to exist only in prokaryotes is the histidine protein kinase family (HPK). HPKs bear little homology with mammalian STKs or PTKs but have distinctive sequence motifs of their own (Davie, J.R. et al. (1995) J. Biol. Chem. 270:19861-19867). A histidine residue in the N-terminal half of the molecule (region I) is an autophosphorylation site. Three additional motifs located in the C-terminal half of the molecule include an invariant asparagine residue in region II and two glycine-rich loops characteristic of nucleotide binding domains in regions III and IV. Recently a branched chain alpha-ketoacid dehydrogenase kinase has been found with characteristics of HPK in rat (Davie et al., supra).

Protein phosphatases regulate the effects of protein kinases by removing phosphate groups from molecules previously activated by kinases. The two principal categories of protein phosphatases are the protein (serine/threonine) phosphatases (PPs) and the protein tyrosine phosphatases (PTPs). PPs dephosphorylate phosphoserine/threonine residues and are important regulators of many cAMP-mediated hormone responses (Cohen, P. (1989) Annu. Rev. Biochem. 58:453-508). PTPs reverse the effects of protein tyrosine kinases and play a significant role in cell cycle and cell signaling

processes (Charbonneau and Tonks, supra). As previously noted, many PTKs are encoded by oncogenes, and oncogenesis is often accompanied by increased tyrosine phosphorylation activity. It is therefore possible that PTPs may prevent or reverse cell transformation and the growth of various cancers by controlling the levels of tyrosine phosphorylation in cells. This hypothesis is supported by studies showing that overexpression of PTPs can suppress transformation in cells, and that specific inhibition of PTPs can enhance cell transformation (Charbonneau and Tonks, supra).

Phospholipid and Inositol-phosphate Signaling

Inositol phospholipids (phosphoinositides) are involved in an intracellular signaling pathway that begins with binding of a signaling molecule to a G-protein linked receptor in the plasma membrane. This leads to the phosphorylation of phosphatidylinositol (PI) residues on the inner side of the plasma membrane to the biphosphate state (PIP_2) by inositol kinases. Simultaneously, the G-protein linked receptor binding stimulates a trimeric G-protein which in turn activates a phosphoinositide-specific phospholipase C-β. Phospholipase C-β then cleaves PIP_2 into two products, inositol triphosphate (IP_3) and diacylglycerol. These two products act as mediators for separate signaling events. IP_3 diffuses through the plasma membrane to induce calcium release from the endoplasmic reticulum (ER), while diacylglycerol remains in the membrane and helps activate protein kinase C, a serine-threonine kinase that phosphorylates selected proteins in the target cell. The calcium response initiated by IP_3 is terminated by the dephosphorylation of IP_3 by specific inositol phosphatases. Cellular responses that are mediated by this pathway are glycogen breakdown in the liver in response to vasopressin, smooth muscle contraction in response to acetylcholine, and thrombin-induced platelet aggregation.

Inositol-phosphate signaling controls tubby, a membrane bound transcriptional regulator that serves as an intracellular messenger of $G\alpha_q$-coupled receptors (Santagata et al. (2001) Science 292:2041-2050). Members of the tubby family contain a C-terminal tubby domain of about 260 amino acids that binds to double-stranded DNA and an N-terminal transcriptional activation domain. Tubby binds to phosphatidylinositol 4,5-bisphosphate, which localizes tubby to the plasma membrane. Activation of the G-protein α_q leads to activation of phospholipase C-β and hydrolysis of phosphoinositide. Loss of phosphatidylinositol 4,5-bisphosphate causes tubby to dissociate from the plasma membrane and to translocate to the nucleus where tubby regulates transcription of its target genes. Defects in the tubby gene are associated with obesity, retinal degeneration, and hearing loss (Boggon, T.J. et al. (1999) Science 286:2119-2125).

Cyclic Nucleotide Signaling

Cyclic nucleotides (cAMP and cGMP) function as intracellular second messengers to transduce a variety of extracellular signals including hormones, light, and neurotransmitters. In

particular, cyclic-AMP dependent protein kinases (PKA) are thought to account for all of the effects of cAMP in most mammalian cells, including various hormone-induced cellular responses. Visual excitation and the phototransmission of light signals in the eye is controlled by cyclic-GMP regulated, Ca^{2+}-specific channels. Because of the importance of cellular levels of cyclic nucleotides in mediating these various responses, regulating the synthesis and breakdown of cyclic nucleotides is an important matter. Thus adenylyl cyclase, which synthesizes cAMP from AMP, is activated to increase cAMP levels in muscle by binding of adrenaline to β-adrenergic receptors, while activation of guanylate cyclase and increased cGMP levels in photoreceptors leads to reopening of the Ca^{2+}-specific channels and recovery of the dark state in the eye. There are nine known transmembrane isoforms of mammalian adenylyl cyclase, as well as a soluble form preferentially expressed in testis. Soluble adenylyl cyclase contains a P-loop, or nucleotide binding domain, and may be involved in male fertility (Buck, J. et al. (1999) Proc. Natl. Acad. Sci. USA 96:79-84).

In contrast, hydrolysis of cyclic nucleotides by cAMP and cGMP-specific phosphodiesterases (PDEs) produces the opposite of these and other effects mediated by increased cyclic nucleotide levels. PDEs appear to be particularly important in the regulation of cyclic nucleotides, considering the diversity found in this family of proteins. At least seven families of mammalian PDEs (PDE1-7) have been identified based on substrate specificity and affinity, sensitivity to cofactors, and sensitivity to inhibitory drugs (Beavo, J.A. (1995) Physiol. Rev. 75:725-748). PDE inhibitors have been found to be particularly useful in treating various clinical disorders. Rolipram, a specific inhibitor of PDE4, has been used in the treatment of depression, and similar inhibitors are undergoing evaluation as anti-inflammatory agents. Theophylline is a nonspecific PDE inhibitor used in the treatment of bronchial asthma and other respiratory diseases (Banner, K.H. and C.P. Page (1995) Eur. Respir. J. 8:996-1000).

Calcium Signaling Molecules

Ca^{2+} is another second messenger molecule that is even more widely used as an intracellular mediator than cAMP. Ca^{2+} can enter the cytosol by two pathways, in response to extracellular signals. One pathway acts primarily in nerve signal transduction where Ca^{2+} enters a nerve terminal through a voltage-gated Ca^{2+} channel. The second is a more ubiquitous pathway in which Ca^{2+} is released from the ER into the cytosol in response to binding of an extracellular signaling molecule to a receptor. Ca^{2+} directly activates regulatory enzymes, such as protein kinase C, which trigger signal transduction pathways. Ca^{2+} also binds to specific Ca^{2+}-binding proteins (CBPs) such as calmodulin (CaM) which then activate multiple target proteins in the cell including enzymes, membrane transport pumps, and ion channels. CaM interactions are involved in a multitude of cellular processes including, but not limited to, gene regulation, DNA synthesis, cell cycle progression, mitosis, cytokinesis,

cytoskeletal organization, muscle contraction, signal transduction, ion homeostasis, exocytosis, and metabolic regulation (Celio, M.R. et al. (1996) Guidebook to Calcium-binding Proteins, Oxford University Press, Oxford, UK, pp. 15-20). Some Ca^{2+} binding proteins are characterized by the presence of one or more EF-hand Ca^{2+} binding motifs, which are comprised of 12 amino acids flanked by α-helices (Celio, supra). The regulation of CBPs has implications for the control of a variety of disorders. Calcineurin, a CaM-regulated protein phosphatase, is a target for inhibition by the immunosuppressive agents cyclosporin and FK506. This indicates the importance of calcineurin and CaM in the immune response and immune disorders (Schwaninger M. et al. (1993) J. Biol Chem. 268:23111-23115). The level of CaM is increased several-fold in tumors and tumor-derived cell lines for various types of cancer (Rasmussen, C.D. and A.R. Means (1989) Trends Neurosci. 12:433-438).

The annexins are a family of calcium-binding proteins that associate with the cell membrane (Towle, C.A. and B.V. Treadwell (1992) J. Biol. Chem. 267:5416-5423). Annexins reversibly bind to negatively charged phospholipids (phosphatidylcholine and phosphatidylserine) in a calcium dependent manner. Annexins participate in various processes pertaining to signal transduction at the plasma membrane, including membrane-cytoskeleton interactions, phospholipase inhibition, anticoagulation, and membrane fusion. Annexins contain four to eight repeated segments of about 60 residues. Each repeat folds into five alpha helices wound into a right-handed superhelix.

G-Protein Signaling

Guanine nucleotide binding proteins (G-proteins) are critical mediators of signal transduction between a particular class of extracellular receptors, the G-protein coupled receptors (GPCRs), and intracellular second messengers such as cAMP and Ca^{2+}. G-proteins are linked to the cytosolic side of a GPCR such that activation of the GPCR by ligand binding stimulates binding of the G-protein to GTP, inducing an "active" state in the G-protein. In the active state, the G-protein acts as a signal to trigger other events in the cell such as the increase of cAMP levels or the release of Ca^{2+} into the cytosol from the ER, which, in turn, regulate phosphorylation and activation of other intracellular proteins. Recycling of the G-protein to the inactive state involves hydrolysis of the bound GTP to GDP by a GTPase activity in the G-protein. (See Alberts, B. et al. (1994) Molecular Biology of the Cell Garland Publishing, Inc. New York, NY, pp.734-759.) The superfamily of G-proteins consists of several families which may be grouped as translational factors, heterotrimeric G-proteins involved in transmembrane signaling processes, and low molecular weight (LMW) G-proteins including the proto-oncogene Ras proteins and products of rab, rap, rho, rac, smg21, smg25, YPT, SEC4, and ARF genes, and tubulins (Kaziro, Y. et al. (1991) Annu. Rev. Biochem. 60:349-400). In all cases, the GTPase activity is regulated through interactions with other proteins.

Heterotrimeric G-proteins are composed of 3 subunits, α, β, and γ, which in their inactive

conformation associate as a trimer at the inner face of the plasma membrane. Gα binds GDP or GTP and contains the GTPase activity. The βγ complex enhances binding of Gα to a receptor. Gγ is necessary for the folding and activity of Gβ (Neer, E.J. et al. (1994) Nature 371:297-300). Multiple homologs of each subunit have been identified in mammalian tissues, and different combinations of subunits have specific functions and tissue specificities (Spiegel, A.M. (1997) J. Inher. Metab. Dis. 20:113-121).

The alpha subunits of heterotrimeric G-proteins can be divided into four distinct classes. The α-s class is sensitive to ADP-ribosylation by pertussis toxin which uncouples the receptor:G-protein interaction. This uncoupling blocks signal transduction to receptors that decrease cAMP levels which normally regulate ion channels and activate phospholipases. The inhibitory α-I class is also susceptible to modification by pertussis toxin which prevents α-I from lowering cAMP levels. Two novel classes of α subunits refractory to pertussis toxin modification are α-q, which activates phospholipase C, and α-12, which has sequence homology with the Drosophila gene concertina and may contribute to the regulation of embryonic development (Simon, M.I. (1991) Science 252:802-808).

The mammalian Gβ and Gγ subunits, each about 340 amino acids long, share more than 80% homology. The Gβ subunit (also called transducin) contains seven repeating units, each about 43 amino acids long. The activity of both subunits may be regulated by other proteins such as calmodulin and phosducin or the neural protein GAP 43 (Clapham, D. and E. Neer (1993) Nature 365:403-406). The β and γ subunits are tightly associated. The β subunit sequences are highly conserved between species, implying that they perform a fundamentally important role in the organization and function of G-protein linked systems (Van der Voorn, L. (1992) FEBS Lett. 307:131-134). They contain seven tandem repeats of the WD-repeat sequence motif, a motif found in many proteins with regulatory functions. WD-repeat proteins contain from four to eight copies of a loosely conserved repeat of approximately 40 amino acids which participates in protein-protein interactions. Mutations and variant expression of β transducin proteins are linked with various disorders. Mutations in LIS1, a subunit of the human platelet activating factor acetylhydrolase, cause Miller-Dieker lissencephaly. RACK1 binds activated protein kinase C, and RbAp48 binds retinoblastoma protein. CstF is required for polyadenylation of mammalian pre-mRNA in vitro and associates with subunits of cleavage-stimulating factor. Defects in the regulation of β-catenin contribute to the neoplastic transformation of human cells. The WD40 repeats of the human F-box protein bTrCP mediate binding to β-catenin, thus regulating the targeted degradation of β-catenin by ubiquitin ligase (Neer, supra; Hart, M. et al. (1999) Curr. Biol. 9:207-210). The γ subunit primary structures are more variable than those of the β subunits. They are often post-translationally modified by isoprenylation and carboxyl-methylation of a

cysteine residue four amino acids from the C-terminus; this appears to be necessary for the interaction of the βγ subunit with the membrane and with other G-proteins. The βγ subunit has been shown to modulate the activity of isoforms of adenylyl cyclase, phospholipase C, and some ion channels. It is involved in receptor phosphorylation via specific kinases, and has been implicated in the p21ras-dependent activation of the MAP kinase cascade and the recognition of specific receptors by G-proteins (Clapham and Neer, supra).

G-proteins interact with a variety of effectors including adenylyl cyclase (Clapham and Neer, supra). The signaling pathway mediated by cAMP is mitogenic in hormone-dependent endocrine tissues such as adrenal cortex, thyroid, ovary, pituitary, and testes. Cancers in these tissues have been related to a mutationally activated form of a Gα, known as the gsp (Gs protein) oncogene (Dhanasekaran, N. et al. (1998) Oncogene 17:1383-1394). Another effector is phosducin, a retinal phosphoprotein, which forms a specific complex with retinal Gβ and Gγ (Gβγ) and modulates the ability of Gβγ to interact with retinal Gα (Clapham and Neer, supra).

Irregularities in the G-protein signaling cascade may result in abnormal activation of leukocytes and lymphocytes, leading to the tissue damage and destruction seen in many inflammatory and autoimmune diseases such as rheumatoid arthritis, biliary cirrhosis, hemolytic anemia, lupus erythematosus, and thyroiditis. Abnormal cell proliferation, including cyclic AMP stimulation of brain, thyroid, adrenal, and gonadal tissue proliferation is regulated by G proteins. Mutations in Gα subunits have been found in growth-hormone-secreting pituitary somatotroph tumors, hyperfunctioning thyroid adenomas, and ovarian and adrenal neoplasms (Meij, J.T.A. (1996) Mol. Cell Biochem. 157:31-38; Aussel, C. et al. (1988) J. Immunol. 140:215-220).

LMW G-proteins are GTPases which regulate cell growth, cell cycle control, protein secretion, and intracellular vesicle interaction. They consist of single polypeptides which, like the alpha subunit of the heterotrimeric G-proteins, are able to bind to and hydrolyze GTP, thus cycling between an inactive and an active state. LMW G-proteins respond to extracellular signals from receptors and activating proteins by transducing mitogenic signals involved in various cell functions. The binding and hydrolysis of GTP regulates the response of LMW G-proteins and acts as an energy source during this process (Bokoch, G.M. and C.J. Der (1993) FASEB J. 7:750-759).

At least sixty members of the LMW G-protein superfamily have been identified and are currently grouped into the ras, rho, arf, sar1, ran, and rab subfamilies. Activated ras genes were initially found in human cancers, and subsequent studies confirmed that ras function is critical in determining whether cells continue to grow or become differentiated. Ras1 and Ras2 proteins stimulate adenylate cyclase (Kaziro, supra), affecting a broad array of cellular processes. Stimulation

of cell surface receptors activates Ras which, in turn, activates cytoplasmic kinases. These kinases translocate to the nucleus and activate key transcription factors that control gene expression and protein synthesis (Barbacid, M. (1987) Annu. Rev. Biochem. 56:779-827, Treisman, R. (1994) Curr. Opin. Genet. Dev. 4:96-98). Other members of the LMW G-protein superfamily have roles in signal transduction that vary with the function of the activated genes and the locations of the G-proteins that initiate the activity. Rho G-proteins control signal transduction pathways that link growth factor receptors to actin polymerization, which is necessary for normal cellular growth and division. The rab, arf, and sar1 families of proteins control the translocation of vesicles to and from membranes for protein processing, localization, and secretion. Vesicle- and target- specific identifiers (v-SNAREs and t-SNAREs) bind to each other and dock the vesicle to the acceptor membrane. The budding process is regulated by the closely related ADP ribosylation factors (ARFs) and SAR proteins, while rab proteins allow assembly of SNARE complexes and may play a role in removal of defective complexes (Rothman, J. and F. Wieland (1996) Science 272:227-234). Ran G-proteins are located in the nucleus of cells and have a key role in nuclear protein import, the control of DNA synthesis, and cell-cycle progression (Hall, A. (1990) Science 249:635-640; Barbacid, M. (1987) Annu. Rev. Biochem. 56:779-827; Ktistakis, N. (1998) BioEssays 20:495-504; and Sasaki, T. and Y. Takai (1998) Biochem. Biophys. Res. Commun. 245:641-645).

Rab proteins have a highly variable amino terminus containing membrane-specific signal information and a prenylated carboxy terminus which determines the target membrane to which the Rab proteins anchor. More than 30 Rab proteins have been identified in a variety of species, and each has a characteristic intracellular location and distinct transport function. In particular, Rab1 and Rab2 are important in ER-to-Golgi transport; Rab3 transports secretory vesicles to the extracellular membrane; Rab5 is localized to endosomes and regulates the fusion of early endosomes into late endosomes; Rab6 is specific to the Golgi apparatus and regulates intra-Golgi transport events; Rab7 and Rab9 stimulate the fusion of late endosomes and Golgi vesicles with lysosomes, respectively; and Rab10 mediates vesicle fusion from the medial Golgi to the trans Golgi. Mutant forms of Rab proteins are able to block protein transport along a given pathway or alter the sizes of entire organelles. Therefore, Rabs play key regulatory roles in membrane trafficking (Schimmöller, I.S. and S.R. Pfeffer (1998) J. Biol. Chem. 243:22161-22164).

The function of Rab proteins in vesicular transport requires the cooperation of many other proteins. Specifically, the membrane-targeting process is assisted by a series of escort proteins (Khosravi-Far, R. et al. (1991) Proc. Natl. Acad. Sci. USA 88:6264-6268). In the medial Golgi, it has been shown that GTP-bound Rab proteins initiate the binding of VAMP-like proteins of the transport vesicle to syntaxin-like proteins on the acceptor membrane, which subsequently triggers a cascade of

protein-binding and membrane-fusion events. After transport, GTPase-activating proteins (GAPs) in the target membrane are responsible for converting the GTP-bound Rab proteins to their GDP-bound state. And finally, guanine-nucleotide dissociation inhibitor (GDI) recruits the GDP-bound proteins to their membrane of origin.

The cycling of LMW G-proteins between the GTP-bound active form and the GDP-bound inactive form is regulated by a variety of proteins. Guanosine nucleotide exchange factors (GEFs) increase the rate of nucleotide dissociation by several orders of magnitude, thus facilitating release of GDP and loading with GTP. The best characterized is the mammalian homolog of the Drosophila Son-of-Sevenless protein. Certain Ras-family proteins are also regulated by guanine nucleotide dissociation inhibitors (GDIs), which inhibit GDP dissociation. The intrinsic rate of GTP hydrolysis of the LMW G-proteins is typically very slow, but it can be stimulated by several orders of magnitude by GAPs (Geyer, M. and A. Wittinghofer (1997) Curr. Opin. Struct. Biol. 7:786-792). Both GEF and GAP activity may be controlled in response to extracellular stimuli and modulated by accessory proteins such as RalBP1 and POB1. Mutant Ras-family proteins, which bind but cannot hydrolyze GTP, are permanently activated, and cause cell proliferation or cancer, as do GEFs that inappropriately activate LMW G-proteins, such as the human oncogene NET1, a Rho-GEF (Drivas, G.T. et al. (1990) Mol. Cell Biol. 10:1793-1798; Alberts, A.S. and R. Treisman (1998) EMBO J. 14:4075-4085).

A member of the ARF family of G-proteins is centaurin beta 1A, a regulator of membrane traffic and the actin cytoskeleton. The centaurin β family of GTPase-activating proteins (GAPs) and Arf guanine nucleotide exchange factors contain pleckstrin homology (PH) domains which are activated by phosphoinositides. PH domains bind phosphoinositides, implicating PH domains in signaling processes. Phosphoinositides have a role in converting Arf-GTP to Arf-GDP via the centaurin β family and a role in Arf activation (Kam, J.L. et al. (2000) J. Biol. Chem. 275:9653-9663). The rho GAP family is also implicated in the regulation of actin polymerization at the plasma membrane and in several cellular processes. The gene ARHGAP6 encodes GTPase-activating protein 6 isoform 4. Mutations in ARHGAP6, seen as a deletion of a 500 kb critical region in Xp22.3, causes the syndrome microphthalmia with linear skin defects (MLS). MLS is an X-linked dominant, male-lethal syndrome (Prakash, S.K. et al. (2000) Hum. Mol. Genet. 9:477-488).

A member of the Rho family of G-proteins is CDC42, a regulator of cytoskeletal rearrangements required for cell division. CDC42 is inactivated by a specific GAP (CDC42GAP) that strongly stimulates the GTPase activity of CDC42 while having a much lesser effect on other Rho family members. CDC42GAP also contains an SH3-binding domain that interacts with the SH3 domains of cell signaling proteins such as p85 alpha and c-Src, suggesting that CDC42GAP may serve

as a link between CDC42 and other cell signaling pathways (Barfod, E.T. et al. (1993) J. Biol. Chem. 268:26059-26062).

The Dbl proteins are a family of GEFs for the Rho and Ras G-proteins (Whitehead, I.P. et al. (1997) Biochim. Biophys. Acta 1332:F1-F23). All Dbl family members contain a Dbl homology (DH) domain of approximately 180 amino acids, as well as a pleckstrin homology (PH) domain located immediately C-terminal to the DH domain. Most Dbl proteins have oncogenic activity, as demonstrated by the ability to transform various cell lines, consistent with roles as regulators of Rho-mediated oncogenic signaling pathways. The kalirin proteins are neuron-specific members of the Dbl family, which are located to distinct subcellular regions of cultured neurons (Johnson, R.C. (2000) J. Cell Biol. 275:19324-19333).

Other regulators of G-protein signaling (RGS) also exist that act primarily by negatively regulating the G-protein pathway by an unknown mechanism (Druey, K.M. et al. (1996) Nature 379:742-746). Some 15 members of the RGS family have been identified. RGS family members are related structurally through similarities in an approximately 120 amino acid region termed the RGS domain and functionally by their ability to inhibit the interleukin (cytokine) induction of MAP kinase in cultured mammalian 293T cells (Druey et al., supra).

The Immuno-associated nucleotide (IAN) family of proteins has GTP-binding activity as indicated by the conserved ATP/GTP-binding site P-loop motif. The IAN family includes IAN-1, IAN-4, IAP38, and IAG-1. IAN-1 is expressed in the immune system, specifically in T cells and thymocytes. Its expression is induced during thymic events (Poirier, G.M.C. et al. (1999) J. Immunol. 163:4960-4969). IAP38 is expressed in B cells and macrophages and its expression is induced in splenocytes by pathogens. IAG-1, which is a plant molecule, is induced upon bacterial infection (Krucken, J. et al. (1997) Biochem. Biophys. Res. Commun. 230:167-170). IAN-4 is a mitochondrial membrane protein which is preferentially expressed in hematopoietic precursor 32D cells transfected with wild-type versus mutant forms of the bcr/abl oncogene. The bcr/abl oncogene is known to be associated with chronic myelogenous leukemia, a clonal myelo-proliferative disorder, which is due to the translocation between the bcr gene on chromosome 22 and the abl gene on chromosome 9. Bcr is the breakpoint cluster region gene and abl is the cellular homolog of the transforming gene of the Abelson murine leukemia virus. Therefore, the IAN family of proteins appears to play a role in cell survival in immune responses and cellular transformation (Daheron, L. et al. (2001) Nucleic Acids Res. 29:1308-1316).

Formin-related genes (FRL) comprise a large family of morphoregulatory genes and have been shown to play important roles in morphogenesis, embryogenesis, cell polarity, cell migration, and cytokinesis through their interaction with Rho family small GTPases. Formin was first identified in

mouse limb deformity (ld) mutants where the distal bones and digits of all limbs are fused and reduced in size. FRL contains formin homology domains FH1, FH2, and FH3. The FH1 domain has been shown to bind the Src homology 3 (SH3) domain, WWP/WW domains, and profilin. The FH2 domain is conserved and was shown to be essential for formin function as disruption at the FH2 domain results in the characteristic *ld* phenotype. The FH3 domain is located at the N-terminus of FRL, and is required for associating with Rac, a Rho family GTPase (Yayoshi-Yamamoto, S. et al. (2000) Mol. Cell. Biol. 20:6872-6881).

Signaling Complex Protein Domains

PDZ domains were named for three proteins in which this domain was initially discovered. These proteins include PSD-95 (postsynaptic density 95), Dlg (Drosophila lethal(1)discs large-1), and ZO-1 (zonula occludens-1). These proteins play important roles in neuronal synaptic transmission, tumor suppression, and cell junction formation, respectively. Since the discovery of these proteins, over sixty additional PDZ-containing proteins have been identified in diverse prokaryotic and eukaryotic organisms. This domain has been implicated in receptor and ion channel clustering and in the targeting of multiprotein signaling complexes to specialized functional regions of the cytosolic face of the plasma membrane. (For a review of PDZ domain-containing proteins, see Ponting, C.P. et al. (1997) Bioessays 19:469-479.) A large proportion of PDZ domains are found in the eukaryotic MAGUK (membrane-associated guanylate kinase) protein family, members of which bind to the intracellular domains of receptors and channels. However, PDZ domains are also found in diverse membrane-localized proteins such as protein tyrosine phosphatases, serine/threonine kinases, G-protein cofactors, and synapse-associated proteins such as syntrophins and neuronal nitric oxide synthase (nNOS). Generally, about one to three PDZ domains are found in a given protein, although up to nine PDZ domains have been identified in a single protein. The glutamate receptor interacting protein (GRIP) contains seven PDZ domains. GRIP is an adaptor that links certain glutamate receptors to other proteins and may be responsible for the clustering of these receptors at excitatory synapses in the brain (Dong, H. et al. (1997) Nature 386:279-284). The Drosophila scribble (SCRIB) protein contains both multiple PDZ domains and leucine-rich repeats. SCRIB is located at the epithelial septate junction, which is analogous to the vertebrate tight junction, at the boundary of the apical and basolateral cell surface. SCRIB is involved in the distribution of apical proteins and correct placement of adherens junctions to the basolateral cell surface (Bilder, D. and N. Perrimon (2000) Nature 403:676-680).

The PX domain is an example of a domain specialized for promoting protein-protein interactions. The PX domain is found in sorting nexins and in a variety of other proteins, including the

PhoX components of NADPH oxidase and the Cpk class of phosphatidylinositol 3-kinase. Most PX domains contain a polyproline motif which is characteristic of SH3 domain-binding proteins (Ponting, C.P. (1996) Protein Sci. 5:2353-2357). SH3 domain-mediated interactions involving the PhoX components of NADPH oxidase play a role in the formation of the NADPH oxidase multi-protein complex (Leto, T.L. et al. (1994) Proc. Natl. Acad. Sci. USA 91:10650-10654; Wilson, L. et al. (1997) Inflamm. Res. 46:265-271).

The SH3 domain is defined by homology to a region of the proto-oncogene c-Src, a cytoplasmic protein tyrosine kinase. SH3 is a small domain of 50 to 60 amino acids that interacts with proline-rich ligands. SH3 domains are found in a variety of eukaryotic proteins involved in signal transduction, cell polarization, and membrane-cytoskeleton interactions. In some cases, SH3 domain-containing proteins interact directly with receptor tyrosine kinases. For example, the SLAP-130 protein is a substrate of the T-cell receptor (TCR) stimulated protein kinase. SLAP-130 interacts via its SH3 domain with the protein SLP-76 to affect the TCR-induced expression of interleukin-2 (Musci, M.A. et al. (1997) J. Biol. Chem. 272:11674-11677). Another recently identified SH3 domain protein is macrophage actin-associated tyrosine-phosphorylated protein (MAYP) which is phosphorylated during the response of macrophages to colony stimulating factor-1 (CSF-1) and is likely to play a role in regulating the CSF-1-induced reorganization of the actin cytoskeleton (Yeung, Y.-G. et al. (1998) J. Biol. Chem. 273:30638-30642). The structure of the SH3 domain is characterized by two antiparallel beta sheets packed against each other at right angles. This packing forms a hydrophobic pocket lined with residues that are highly conserved between different SH3 domains. This pocket makes critical hydrophobic contacts with proline residues in the ligand (Feng, S. et al. (1994) Science 266:1241-1247).

A novel domain, called the WW domain, resembles the SH3 domain in its ability to bind proline-rich ligands. This domain was originally discovered in dystrophin, a cytoskeletal protein with direct involvement in Duchenne muscular dystrophy (Bork, P. and M. Sudol (1994) Trends Biochem. Sci. 19:531-533). WW domains have since been discovered in a variety of intracellular signaling molecules involved in development, cell differentiation, and cell proliferation. The structure of the WW domain is composed of beta strands grouped around four conserved aromatic residues, generally tryptophan.

Like SH3, the SH2 domain is defined by homology to a region of c-Src. SH2 domains interact directly with phospho-tyrosine residues, thus providing an immediate mechanism for the regulation and transduction of receptor tyrosine kinase-mediated signaling pathways. For example, as many as ten distinct SH2 domains are capable of binding to phosphorylated tyrosine residues in the activated PDGF receptor, thereby providing a highly coordinated and finely tuned response to ligand-mediated receptor

activation. (Reviewed in Schaffhausen, B. (1995) Biochim. Biophys. Acta. 1242:61-75.) The BLNK protein is a linker protein involved in B cell activation, that bridges B cell receptor-associated kinases with SH2 domain effectors that link to various signaling pathways (Fu, C. et al. (1998) Immunity 9:93-103).

The pleckstrin homology (PH) domain was originally identified in pleckstrin, the predominant substrate for protein kinase C in platelets. Since its discovery, this domain has been identified in over 90 proteins involved in intracellular signaling or cytoskeletal organization. Proteins containing the pleckstrin homology domain include a variety of kinases, phospholipase-C isoforms, guanine nucleotide release factors, and GTPase activating proteins. For example, members of the FGD1 family contain both Rho-guanine nucleotide exchange factor (GEF) and PH domains, as well as a FYVE zinc finger domain. FGD1 is the gene responsible for faciogenital dysplasia, an inherited skeletal dysplasia (Pasteris, N.G. and J.L. Gorski (1999) Genomics 60:57-66). Many PH domain proteins function in association with the plasma membrane, and this association appears to be mediated by the PH domain itself. PH domains share a common structure composed of two antiparallel beta sheets flanked by an amphipathic alpha helix. Variable loops connecting the component beta strands generally occur within a positively charged environment and may function as ligand binding sites (Lemmon, M.A. et al. (1996) Cell 85:621-624). Ankyrin (ANK) repeats mediate protein-protein interactions associated with diverse intracellular signaling functions. For example, ANK repeats are found in proteins involved in cell proliferation such as kinases, kinase inhibitors, tumor suppressors, and cell cycle control proteins. (See, for example, Kalus, W. et al. (1997) FEBS Lett. 401:127-132; Ferrante, A.W. et al. (1995) Proc. Natl. Acad. Sci. USA 92:1911-1915.) These proteins generally contain multiple ANK repeats, each composed of about 33 amino acids. Myotrophin is an ANK repeat protein that plays a key role in the development of cardiac hypertrophy, a contributing factor to many heart diseases. Structural studies show that the myotrophin ANK repeats, like other ANK repeats, each form a helix-turn-helix core preceded by a protruding "tip." These tips are of variable sequence and may play a role in protein-protein interactions. The helix-turn-helix region of the ANK repeats stack on top of one another and are stabilized by hydrophobic interactions (Yang, Y. et al. (1998) Structure 6:619-626). Members of the ASB protein family contain a suppressor of cytokine signaling (SOCS) domain as well as multiple ankyrin repeats (Hilton, D.J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:114-119).

The tetratricopeptide repeat (TPR) is a 34 amino acid repeated motif found in organisms from bacteria to humans. TPRs are predicted to form ampipathic helices, and appear to mediate protein-protein interactions. TPR domains are found in CDC16, CDC23, and CDC27, members of the anaphase promoting complex which targets proteins for degradation at the onset of anaphase. Other processes involving TPR proteins include cell cycle control, transcription repression, stress response,

and protein kinase inhibition (Lamb, J.R. et al. (1995) Trends Biochem. Sci. 20:257-259).

The armadillo/beta-catenin repeat is a 42 amino acid motif which forms a superhelix of alpha helices when tandemly repeated. The structure of the armadillo repeat region from beta-catenin revealed a shallow groove of positive charge on one face of the superhelix, which is a potential binding surface. The armadillo repeats of beta-catenin, plakoglobin, and p120^{cas} bind the cytoplasmic domains of cadherins. Beta-catenin/cadherin complexes are targets of regulatory signals that govern cell adhesion and mobility (Huber, A.H. et al. (1997) Cell 90:871-882).

Eight tandem repeats of about 40 residues (WD-40 repeats), each containing a central Trp-Asp motif, make up beta-transducin (G-beta), which is one of the three subunits (alpha, beta, and gamma) of the guanine nucleotide-binding proteins (G proteins). In higher eukaryotes G-beta exists as a small multigene family of highly conserved proteins of about 340 amino acid residues.

Expression profiling

Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

Breast Cancer

There are more than 180,000 new cases of breast cancer diagnosed each year, and the mortality rate for breast cancer approaches 10% of all deaths in females between the ages of 45-54 (K. Gish (1999) AWIS Magazine 28:7-10). However the survival rate based on early diagnosis of localized breast cancer is extremely high (97%), compared with the advanced stage of the disease in which the tumor has spread beyond the breast (22%). Current procedures for clinical breast examination are lacking in sensitivity and specificity, and efforts are underway to develop comprehensive gene expression profiles for breast cancer that may be used in conjunction with conventional screening methods to improve diagnosis and prognosis of this disease (Perou CM et al. (2000) Nature 406:747-752).

Breast cancer is a genetic disease commonly caused by mutations in cellular disease. Mutations in two genes, BRCA1 and BRCA2, are known to greatly predispose a woman to breast cancer and may be passed on from parents to children (Gish, supra). However, this type of hereditary breast cancer accounts for only about 5% to 9% of breast cancers, while the vast majority of breast

cancer is due to noninherited mutations that occur in breast epithelial cells.

A good deal is already known about the expression of specific genes associated with breast cancer. For example, the relationship between expression of epidermal growth factor (EGF) and its receptor, EGFR, to human mammary carcinoma has been particularly well studied. (See Khazaie et al., supra, and references cited therein for a review of this area.) Overexpression of EGFR, particularly coupled with down-regulation of the estrogen receptor, is a marker of poor prognosis in breast cancer patients. In addition, EGFR expression in breast tumor metastases is frequently elevated relative to the primary tumor, suggesting that EGFR is involved in tumor progression and metastasis. This is supported by accumulating evidence that EGF has effects on cell functions related to metastatic potential, such as cell motility, chemotaxis, secretion and differentiation. Changes in expression of other members of the erbB receptor family, of which EGFR is one, have also been implicated in breast cancer. The abundance of erbB receptors, such as HER-2/neu, HER-3, and HER-4, and their ligands in breast cancer points to their functional importance in the pathogenesis of the disease, and may therefore provide targets for therapy of the disease (Bacus, SS et al. (1994) Am J Clin Pathol 102:S13-S24). Other known markers of breast cancer include a human secreted frizzled protein mRNA that is downregulated in breast tumors; the matrix G1a protein which is overexpressed is human breast carcinoma cells; Drg1 or RTP, a gene whose expression is diminished in colon, breast, and prostate tumors; maspin, a tumor suppressor gene downregulated in invasive breast carcinomas; and CaN19, a member of the S100 protein family, all of which are down regulated in mammary carcinoma cells relative to normal mammary epithelial cells (Zhou Z et al. (1998) Int J Cancer 78:95-99; Chen, L et al. (1990) Oncogene 5:1391-1395; Ulrix W et al (1999) FEBS Lett 455:23-26; Sager, R et al. (1996) Curr Top Microbiol Immunol 213:51-64; and Lee, SW et al. (1992) Proc Natl Acad Sci USA 89:2504-2508).

Cell lines derived from human mammary epithelial cells at various stages of breast cancer provide a useful model to study the process of malignant transformation and tumor progression as it has been shown that these cell lines retain many of the properties of their parental tumors for lengthy culture periods (Wistuba II et al. (1998) Clin Cancer Res 4:2931-2938). Such a model is particularly useful for comparing phenotypic and molecular characteristics of human mammary epithelial cells at various stages of malignant transformation.

Prostate Cancer

Prostate cancer is a common malignancy in men over the age of 50, and the incidence increases with age. In the US, there are approximately 132,000 newly diagnosed cases of prostate cancer and more than 33,000 deaths from the disorder each year.

Once cancer cells arise in the prostate, they are stimulated by testosterone to a more rapid

growth. Thus, removal of the testes can indirectly reduce both rapid growth and metastasis of the cancer. Over 95 percent of prostatic cancers are adenocarcinomas which originate in the prostatic acini. The remaining 5 percent are divided between squamous cell and transitional cell carcinomas, both of which arise in the prostatic ducts or other parts of the prostate gland.

As with most cancers, prostate cancer develops through a multistage progression ultimately resulting in an aggressive, metastatic phenotype. The initial step in tumor progression involves the hyperproliferation of normal luminal and/or basal epithelial cells that become hyperplastic and evolve into early-stage tumors. The early-stage tumors are localized in the prostate but eventually may metastasize, particularly to the bone, brain or lung. About 80% of these tumors remain responsive to androgen treatment, an important hormone controlling the growth of prostate epithelial cells. However, in its most advanced state, cancer growth becomes androgen-independent and there is currently no known treatment for this condition.

A primary diagnostic marker for prostate cancer is prostate specific antigen (PSA). PSA is a tissue-specific serine protease almost exclusively produced by prostatic epithelial cells. The quantity of PSA correlates with the number and volume of the prostatic epithelial cells, and consequently, the levels of PSA are an excellent indicator of abnormal prostate growth. Men with prostate cancer exhibit an early linear increase in PSA levels followed by an exponential increase prior to diagnosis. However, since PSA levels are also influenced by factors such as inflammation, androgen and other growth factors, some scientists maintain that changes in PSA levels are not useful in detecting individual cases of prostate cancer.

Current areas of cancer research provide additional prospects for markers as well as potential therapeutic targets for prostate cancer. Several growth factors have been shown to play a critical role in tumor development, growth, and progression. The growth factors Epidermal Growth Factor (EGF), Fibroblast Growth Factor (FGF), and Tumor Growth Factor alpha (TGFα) are important in the growth of normal as well as hyperproliferative prostate epithelial cells, particularly at early stages of tumor development and progression, and affect signaling pathways in these cells in various ways (Lin J et al. (1999) Cancer Res. 59:2891-2897; Putz T et al. (1999) Cancer Res 59:227-233). The TGF-β family of growth factors are generally expressed at increased levels in human cancers and the high expression levels in many cases correlates with advanced stages of malignancy and poor survival (Gold LI (1999) Crit Rev Oncog 10:303-360). Finally, there are human cell lines representing both the androgen-dependent stage of prostate cancer (LNCap) as well as the androgen-independent, hormone refractory stage of the disease (PC3 and DU-145) that have proved useful in studying gene expression patterns associated with the progression of prostate cancer, and the effects of cell treatments on these expressed genes (Chung TD (1999) Prostate 15:199-207).

The discovery of new molecules for disease detection and treatment, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, autoimmune/inflammatory, developmental, and neurological disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of molecules for disease detection and treatment.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, molecules for disease detection and treatment, referred to collectively as "MDDT" and individually as "MDDT-1," "MDDT-2," "MDDT-3," "MDDT-4," "MDDT-5," "MDDT-6," "MDDT-7," "MDDT-8," "MDDT-9," "MDDT-10," "MDDT-11," "MDDT-12," "MDDT-13," "MDDT-14," "MDDT-15," "MDDT-16," "MDDT-17," "MDDT-18," "MDDT-19," "MDDT-20," "MDDT-21," "MDDT-22," "MDDT-23," "MDDT-24," "MDDT-25," "MDDT-26," "MDDT-27," "MDDT-28," "MDDT-29," "MDDT-30," "MDDT-31," "MDDT-32," "MDDT-33," "MDDT-34," "MDDT-35," "MDDT-36," "MDDT-37," "MDDT-38," and "MDDT-39." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-39. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-39.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-39. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-39. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:40-78.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group

consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-39. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-39. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-39.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:40-78, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:40-78, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample,

said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:40-78, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:40-78, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:40-78, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:40-78, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-39. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional MDDT, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an

agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-39. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional MDDT, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-39. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional MDDT, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-39. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test

compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-39. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:40-78, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:40-78, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:40-78, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:40-78, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a

polynucleotide sequence selected from the group consisting of SEQ ID NO:40-78, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

Table 8 shows single nucleotide polymorphisms found in polynucleotide sequences of the invention, along with allele frequencies in different human populations.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will

be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"MDDT" refers to the amino acid sequences of substantially purified MDDT obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of MDDT. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of MDDT either by directly interacting with MDDT or by acting on components of the biological pathway in which MDDT participates.

An "allelic variant" is an alternative form of the gene encoding MDDT. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding MDDT include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as MDDT or a polypeptide with at least one functional characteristic of MDDT. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding MDDT, and improper or unexpected hybridization to allelic variants, with

a locus other than the normal chromosomal locus for the polynucleotide sequence encoding MDDT. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent MDDT. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of MDDT is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of MDDT. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of MDDT either by directly interacting with MDDT or by acting on components of the biological pathway in which MDDT participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind MDDT polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to

immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic MDDT, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding MDDT or fragments of MDDT may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His

Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of MDDT or the polynucleotide encoding MDDT which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a

fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:40-78 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:40-78, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:40-78 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:40-78 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:40-78 and the region of SEQ ID NO:40-78 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-39 is encoded by a fragment of SEQ ID NO:40-78. A fragment of SEQ ID NO:1-39 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-39. For example, a fragment of SEQ ID NO:1-39 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-39. The precise length of a fragment of SEQ ID NO:1-39 and the region of SEQ ID NO:1-39 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default

parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62
Reward for match: 1
Penalty for mismatch: -2
Open Gap: 5 and Extension Gap: 2 penalties
Gap x drop-off: 50
Expect: 10
Word Size: 11
Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported

by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment

length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as

formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of MDDT which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of MDDT which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of MDDT. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of MDDT.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a

functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an MDDT may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of MDDT.

"Probe" refers to nucleic acid sequences encoding MDDT, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge

MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing MDDT, nucleic acids encoding MDDT, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression

by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In one alternative, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding

polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human molecules for disease detection and treatment (MDDT), the polynucleotides encoding MDDT, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, autoimmune/inflammatory, developmental, and neurological disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown. Column 6 shows the Incyte ID numbers of physical, full length clones corresponding to the polypeptide and polynucleotide sequences of the invention. The full length clones encode polypeptides which have at least 95% sequence identity to the polypeptide sequences shown in column 3.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte

polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are molecules for disease detection and treatment. For example, SEQ ID NO:2 is 53% identical, from residue G3 to residue G172 and A183 to residue G659, to human mitogen inducible gene mig-2 (GenBank ID g505033) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.2 e-197; which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains a pleckstrin homology (PH) domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLAST analyses provide further corroborative evidence that SEQ ID NO:2 is a cell signaling molecule.

In another example, SEQ ID NO:14 is 91% identical, from residue M1 to residue V659, to mouse DMR-N9 (GenBank ID g817954) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:14 also contains WD repeats as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and additional BLAST analyses against the PRODOM and DOMO databases provide further corroborative evidence that SEQ ID NO:14 is a protein associated with myotonic dystrophy.

In another example, SEQ ID NO:24 is 41% identical, from residue I97 to residue N378, to sponge longevity gene SDLAGL (GenBank ID g9798556) as determined by the Basic Local

Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 9.8e-58, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:24 also contains a homeobox domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from additional BLAST analyses provide further corroborative evidence that SEQ ID NO:24 is a longevity assurance gene.

In another example, SEQ ID NO:26 is 75% identical, from residue M1 to residue S1273, to a human protein, ORF2, which contains a reverse transcriptase domain (GenBank ID g339777) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:26 also contains AP endonuclease family and reverse transcriptase domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, and further BLAST analyses provide further corroborative evidence that SEQ ID NO:26 contains a reverse transcriptase domain.

In another example, SEQ ID NO:33 is 90% identical, from residue M1 to residue N1275, to a predicted polypeptide comprising a reverse transcriptase domain (GenBank ID g339771) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:33 also contains a reverse transcriptase domain and an AP endonuclease domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from additional BLAST analysis provide further corroborative evidence that SEQ ID NO:33 is a reverse transcriptase. SEQ ID NO:1, SEQ ID NO:3-13, SEQ ID NO:15-23, SEQ ID NO:25, SEQ ID NO:27-32 and SEQ ID NO:34-39 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-39 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5′) and stop (3′) positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide

sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:40-78 or that distinguish between SEQ ID NO:40-78 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

Table 8 shows single nucleotide polymorphisms (SNPs) found in polynucleotide sequences of the invention, along with allele frequencies in different human populations. Columns 1 and 2 show the polynucleotide sequence identification number (SEQ ID NO:) and the corresponding Incyte project identification number (PID) for polynucleotides of the invention. Column 3 shows the Incyte identification number for the EST in which the SNP was detected (EST ID), and column 4 shows the identification number for the SNP (SNP ID). Column 5 shows the position within the EST sequence at which the SNP is located (EST SNP), and column 6 shows the position of the SNP within the full-length polynucleotide sequence (CB1 SNP). Column 7 shows the allele found in the EST sequence. Columns 8 and 9 show the two alleles found at the SNP site. Column 10 shows the amino acid encoded by the codon including the SNP site, based upon the allele found in the EST. Columns 11-14 show the frequency of allele 1 in four different human populations. An entry of n/d (not detected) indicates that the frequency of allele 1 in the population was too low to be detected, while n/a (not available) indicates that the allele frequency was not determined for the population.

The invention also encompasses MDDT variants. A preferred MDDT variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid

sequence identity to the MDDT amino acid sequence, and which contains at least one functional or structural characteristic of MDDT.

The invention also encompasses polynucleotides which encode MDDT. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:40-78, which encodes MDDT. The polynucleotide sequences of SEQ ID NO:40-78, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding MDDT. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding MDDT. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:40-78 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:40-78. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of MDDT.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding MDDT. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding MDDT, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding MDDT over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding MDDT. For example, a polynucleotide comprising a sequence of SEQ ID NO:78 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:47. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of MDDT.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding MDDT, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide

sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring MDDT, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode MDDT and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring MDDT under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding MDDT or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding MDDT and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode MDDT and MDDT derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding MDDT or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:40-78 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other

systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding MDDT may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the

emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode MDDT may be cloned in recombinant DNA molecules that direct expression of MDDT, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express MDDT.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter MDDT-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of MDDT, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding MDDT may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids

Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, MDDT itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of MDDT, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, *supra*, pp. 28-53.)

In order to express a biologically active MDDT, the nucleotide sequences encoding MDDT or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding MDDT. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding MDDT. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding MDDT and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding MDDT and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995)

Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding MDDT. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, *supra*; Ausubel, *supra*; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding MDDT. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding MDDT can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Invitrogen). Ligation of sequences encoding MDDT into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of MDDT are needed, e.g. for the production of antibodies, vectors which direct high level expression of MDDT may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of MDDT. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such

vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of MDDT. Transcription of sequences encoding MDDT may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding MDDT may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses MDDT in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of MDDT in cell lines is preferred. For example, sequences encoding MDDT can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue

culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding MDDT is inserted within a marker gene sequence, transformed cells containing sequences encoding MDDT can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding MDDT under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding MDDT and that express MDDT may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of MDDT using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on MDDT is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See,

e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding MDDT include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding MDDT, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding MDDT may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode MDDT may be designed to contain signal sequences which direct secretion of MDDT through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding MDDT may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric MDDT protein

containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of MDDT activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the MDDT encoding sequence and the heterologous protein sequence, so that MDDT may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled MDDT may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

MDDT of the present invention or fragments thereof may be used to screen for compounds that specifically bind to MDDT. At least one and up to a plurality of test compounds may be screened for specific binding to MDDT. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules. In one embodiment, the compound thus identified is closely related to the natural ligand of MDDT, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):Chapter 5.) In another embodiment, the compound thus identified is a natural ligand of a receptor MDDT. (See, e.g., Howard, A.D. et al. (2001) Trends Pharmacol. Sci.22:132-140; Wise, A. et al. (2002) Drug Discovery Today 7:235-246.)

In other embodiments, the compound can be closely related to the natural receptor to which MDDT binds, at least a fragment of the receptor, or a fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for MDDT which is capable of propagating a signal, or a decoy receptor for MDDT which is not capable of propagating a signal (Ashkenazi, A. and V.M. Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336). The compound can be rationally designed

using known techniques. Examples of such techniques include those used to construct the compound etanercept (ENBREL; Immunex Corp., Seattle WA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG_1 (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit MDDT involves producing appropriate cells which express MDDT, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing MDDT or cell membrane fractions which contain MDDT are then contacted with a test compound and binding, stimulation, or inhibition of activity of either MDDT or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with MDDT, either in solution or affixed to a solid support, and detecting the binding of MDDT to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No. 6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands. (See, e.g., Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30.) In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to improve or alter its ability to bind to its natural receptors. (See, e.g., Cunningham, B.C. and J.A. Wells (1991) Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. et al. (1991) J. Biol. Chem. 266:10982-10988.)

MDDT of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of MDDT. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for MDDT activity, wherein MDDT is combined with at least one test compound, and the activity of MDDT in the presence of a test compound is compared with the activity of MDDT in the absence of the test compound. A change in the activity of MDDT in the presence of the test compound is indicative of a compound that modulates the activity of MDDT. Alternatively, a test compound is combined with an

in vitro or cell-free system comprising MDDT under conditions suitable for MDDT activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of MDDT may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding MDDT or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams; and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding MDDT may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding MDDT can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding MDDT is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress MDDT, e.g., by secreting MDDT in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between

regions of MDDT and molecules for disease detection and treatment. In addition, examples of tissues expressing MDDT can be found in Table 6 and can also be found in Example XI. Therefore, MDDT appears to play a role in cell proliferative, autoimmune/inflammatory, developmental, and neurological disorders. In the treatment of disorders associated with increased MDDT expression or activity, it is desirable to decrease the expression or activity of MDDT. In the treatment of disorders associated with decreased MDDT expression or activity, it is desirable to increase the expression or activity of MDDT.

Therefore, in one embodiment, MDDT or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of MDDT. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and

neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; and a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia.

In another embodiment, a vector capable of expressing MDDT or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of MDDT including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified MDDT in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of MDDT including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of MDDT may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of MDDT including, but not limited to, those listed above.

In a further embodiment, an antagonist of MDDT may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of MDDT. Examples of such disorders include, but are not limited to, those cell proliferative, autoimmune/inflammatory, developmental, and neurological disorders described above. In one aspect, an antibody which

specifically binds MDDT may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express MDDT.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding MDDT may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of MDDT including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of MDDT may be produced using methods which are generally known in the art. In particular, purified MDDT may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind MDDT. Antibodies to MDDT may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with MDDT or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to MDDT have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches

of MDDT amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to MDDT may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce MDDT-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for MDDT may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between MDDT and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering MDDT epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for MDDT. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of MDDT-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple MDDT epitopes, represents the average affinity, or avidity, of the antibodies for MDDT. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular MDDT epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the MDDT-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of MDDT, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of MDDT-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, *supra*, and Coligan et al. *supra*.)

In another embodiment of the invention, the polynucleotides encoding MDDT, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding MDDT. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding MDDT. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g.,

Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, *supra*; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding MDDT may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in MDDT expression or regulation causes disease, the expression of MDDT from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in MDDT are treated by constructing mammalian expression vectors encoding MDDT and introducing these vectors by mechanical means into MDDT-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem.

62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of MDDT include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). MDDT may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, *supra*)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding MDDT from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to MDDT expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding MDDT under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et

al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding MDDT to cells which have one or more genetic abnormalities with respect to the expression of MDDT. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding MDDT to target cells which have one or more genetic abnormalities with respect to the expression of MDDT. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing MDDT to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple

plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding MDDT to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for MDDT into the alphavirus genome in place of the capsid-coding region results in the production of a large number of MDDT-coding RNAs and the synthesis of high levels of MDDT in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of MDDT into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example,

engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding MDDT.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding MDDT. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding MDDT. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased MDDT expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding MDDT may be therapeutically useful, and in the treatment of disorders associated with

decreased MDDT expression or activity, a compound which specifically promotes expression of the polynucleotide encoding MDDT may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding MDDT is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding MDDT are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding MDDT. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of

such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of MDDT, antibodies to MDDT, and mimetics, agonists, antagonists, or inhibitors of MDDT.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising MDDT or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, MDDT or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for

administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example MDDT or fragments thereof, antibodies of MDDT, and agonists, antagonists or inhibitors of MDDT, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind MDDT may be used for the diagnosis of disorders characterized by expression of MDDT, or in assays to monitor patients being treated with MDDT or agonists, antagonists, or inhibitors of MDDT. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for MDDT include methods which utilize the antibody and a label to detect MDDT in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of

reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring MDDT, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of MDDT expression. Normal or standard values for MDDT expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to MDDT under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of MDDT expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding MDDT may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of MDDT may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of MDDT, and to monitor regulation of MDDT levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding MDDT or closely related molecules may be used to identify nucleic acid sequences which encode MDDT. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding MDDT, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the MDDT encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:40-78 or from genomic sequences including promoters, enhancers, and introns of the MDDT gene.

Means for producing specific hybridization probes for DNAs encoding MDDT include the cloning of polynucleotide sequences encoding MDDT or MDDT derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding MDDT may be used for the diagnosis of disorders

associated with expression of MDDT. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; and a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including

kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia. The polynucleotide sequences encoding MDDT may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered MDDT expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding MDDT may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding MDDT may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding MDDT in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of MDDT, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding MDDT, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding MDDT may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding MDDT, or a fragment of a polynucleotide complementary to the polynucleotide encoding MDDT, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding MDDT may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding MDDT are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the

alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations. (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641.)

Methods which may also be used to quantify the expression of MDDT include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic

profile.

In another embodiment, MDDT, fragments of MDDT, or antibodies specific for MDDT may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for

example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for MDDT to quantify the levels of MDDT expression. In one embodiment, the antibodies are used as elements on a

microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding MDDT may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding MDDT on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, MDDT, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a

solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between MDDT and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with MDDT, or fragments thereof, and washed. Bound MDDT is then detected by methods well known in the art. Purified MDDT can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding MDDT specifically compete with a test compound for binding MDDT. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with MDDT.

In additional embodiments, the nucleotide sequences which encode MDDT may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, in particular U.S. Ser. No. 60/293,723, U.S. Ser. No. 60/295,257, U.S. Ser. No. 60/297,220, U.S. Ser. No. 60/300,526, U.S. Ser. No. 60/301,874 and U.S. Ser. No. 60/359,413 are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with

chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Invitrogen), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Invitrogen), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Invitrogen.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal

cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra; unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from Homo sapiens, Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences.

Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:40-78. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative molecules for disease detection and treatment were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates

predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode molecules for disease detection and treatment, the encoded polypeptides were analyzed by querying against PFAM models for molecules for disease detection and treatment. Potential molecules for disease detection and treatment were also identified by homology to Incyte cDNA sequences that had been annotated as molecules for disease detection and treatment. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data "Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or

genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of MDDT Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:40-78 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:40-78 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid

markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding MDDT are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA

sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding MDDT. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of MDDT Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μl PICOGREEN

quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37 °C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94 °C, 3 min; Step 2: 94 °C, 15 sec; Step 3: 60 °C, 1 min; Step 4: 72 °C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72 °C, 5 min; Step 7: storage at 4 °C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in MDDT Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:40-78 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the

identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezualan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:40-78 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon

membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription

reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85° C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60° C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample

mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and

measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte). Array elements that exhibited at least about a two-fold change in expression, a signal-to-background ratio of at least 2.5, and an element spot size of at least 40% were identified as differentially expressed using the GEMTOOLS program (Incyte Genomics).

Expression

For example, expression of SEQ ID NO:40 was upregulated in PBMC cells stimulated with lipopolysaccharide (LPS), a component of the bacterial cell wall which induces an inflammatory response. PBMCs collected from the blood of four healthy donors was stimulated with 1 μg/ml LPS for 4, 24, and 72 hours. The PBMCs contained about 52% lymphocytes (12% B-cells and 40% T-cells), 20% natural killer cells, 25% monocytes, and 3% various cells including dendritic cells. Stimulated cells were then compared to untreated, time-matched controls. Similarly, expression of SEQ ID NO:40 was upregulated in vascular tissue stimulated with the inflammatory cytokine TNFα or a combination of the protein kinase C activator, PMA, and ionomycin. Cells isolated from vascular smooth muscle, including human coronary artery smooth muscle cells (CASMC), and vascular endothelium, including HUVECs, were grown to 85% confluence in SmGM-2 or EGM-2, respectively, at 37°C, 5% CO_2. Cells were then stimulated with either 10 ng/ml TNFα or 1 μM PMA, 1 μg/ml ionomycin over a defined time course. Upregulation of SEQ ID NO:40 in treated cells relative to untreated, time-matched controls was seen within 1-4 hours following treatment.

Expression of SEQ ID NO:42 was downregulated in ovarian adenocarcinoma and a breast adenocarcinoma cell line, BT-20, relative to normal ovary and breast, respectively. Ovarian tumor tissue obtained from a 79-year-old female was compared to normal ovary obtained from the same donor. BT-20 is a breast adenocarcinoma line derived in vitro from cells emigrating out of thin slices of a tumor mass isolated form a 74-year-old female. BT-20 cells were compared to primary mammary epithelial cells (HMEC) and a breast mammary gland cell line (MCF-10A) isolated from a 36-year-old woman with fibrocystic disease. The breast cell lines were grown in basal medium in the absence of growth factors and hormones for 24 hours prior to the comparison.

Expression of SEQ ID NO:45 was upregulated in THP-1 promonocytes stimulated with PMA and ionomycin. THP-1 is a promonocyte cell line isolated from the peripheral blood of a 1-year-old male with acute monocytic leukemia. Upon stimulation with PMA, THP-1 differentiates into a

macrophage-like cell that displays many characteristics of peripheral human macrophages. THP-1 cells stimulated *in vitro* with 0.1 μM PMA and 1 μg/ml ionomycin for 0.5, 1, 2, 4, and 8 hours were compared to untreated, time-matched control cells. Expression of SEQ ID NO:45 was downregulated in several breast cell cancer lines relative to HMECs. Experiments on breast cell lines were as described above. Cell lines included BT-20, BT474, BT483, Hs578T, MCF-7, and MD-AMB-468.

Expression of SEQ ID NO:48 was upregulated in HUVECs stimulated with TNFα following pre-treatment with either PMA or a low dose of TNFα. HUVECs were pre-treated with either 100 nM PMA or 0.1 ng/ml TNFα for 24 hours, washed, and then stimulated with TNFα for an additional 1, 4, and 24 hours. HUVECs were cultured in IMDM, 10% fetal calf serum at 37°C, 5% CO_2. Treated cells were compared to untreated, time-matched controls.

For example, SEQ ID NO:68 is upregulated 3.4 fold in mature DC versus monocytes, suggesting that SEQ ID NO:68, encoding SEQ ID NO:29, could be used for example, to understand the process by which monocytes differentiate into immature dendritic cells and eventually allow manipulation of the immune system leading to potential immunotherapies for diseases such as cancer, AIDS, and infectious diseases; and enhancing vaccine efficacy.

In another example, SEQ ID NO:78 showed differential expression in inflammatory responses as determined by microarray analysis. The expression of SEQ ID NO:78 was increased by at least two fold in THP-1 human promonocyte line which had been stimulated for 26 hours with 1 μM PMA (phorbol 12-myristate 13-acetate) when compared to untreated THP-1 cells. PMA is a broad activator of the protein kinase C-dependent pathways. THP-1 is promonocyte line derived from peripheral blood of a 1 year old male with acute monocytic leukemia. The cell line acquires monocytic characteristics upon stimulation with PMA. Monocytes play a critical role in the initiation and maintenance of inflammatory immune responses. Therefore, SEQ ID NO:78 is useful in diagnostic assays for inflammatory responses.

Further, as determined by microarray analysis, SEQ ID NO:78 showed differential expression in SKBr3 breast carcinoma cell line versus HMEC primary mammary epithelial cells and MCF10A breast mammary gland cells. SkBR3 is a breast adenocarcinoma cell line isolated from a malignant pleural effusion of a 43-year-old female. HMEC, a primary mammary epithelial cell line was derived from normal human mammary tissue (Clonetics, San Diego, CA). MCF10A, a breast mammary gland (luminal ductal characteristics) cell line was isolated from a 36 year old woman with fibrocystic breast disease. The microarray experiments showed that the expression of SEQ ID NO:78 was increased by at least two fold in SKBr3 breast adenocarcinoma line relative to cells from the primary mammary epithelial cell line, HMEC and the breast mammary gland cell line, MCF10A. Therefore, SEQ ID NO:78 is useful as diagnostic markers or as potential therapeutic targets for breast cancer.

In an alternative example, SEQ ID NO:78 showed differential expression in MDAPCa2b prostate adenocarcinoma cell line versus PrEC normal prostate epithelial cells as determined by microarray analysis. MDAPCa2b is a prostate adenocarcinoma cell line isolated from a metastatic site in the bone of a 63-year-old male. MDAPCa2b cell line expresses prostate specific antigen (PSA) and androgen receptor, grows in vitro and in vivo, and is androgen sensitive. The normal epithelial cell line, PrEC, is a primary prostate epithelial cell line isolated from a normal donor. The experiment showed that the expression of SEQ ID NO:78 was increased by at least two fold in MDAPCa2b cell line relative to PrECs. Therefore, SEQ ID NO:78 is useful as a diagnostic marker or as a potential therapeutic target for prostate cancer.

XII. Complementary Polynucleotides

Sequences complementary to the MDDT-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring MDDT. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of MDDT. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the MDDT-encoding transcript.

XIII. Expression of MDDT

Expression and purification of MDDT is achieved using bacterial or virus-based expression systems. For expression of MDDT in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express MDDT upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of MDDT in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding MDDT by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (*Sf9*) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et

al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, MDDT is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Biosciences). Following purification, the GST moiety can be proteolytically cleaved from MDDT at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified MDDT obtained by these methods can be used directly in the assays shown in Examples XVII and XVIII, where applicable.

XIV. Functional Assays

MDDT function is assessed by expressing the sequences encoding MDDT at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding

of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of MDDT on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding MDDT and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding MDDT and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of MDDT Specific Antibodies

MDDT substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the MDDT amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-MDDT activity by, for example, binding the peptide or MDDT to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring MDDT Using Specific Antibodies

Naturally occurring or recombinant MDDT is substantially purified by immunoaffinity chromatography using antibodies specific for MDDT. An immunoaffinity column is constructed by covalently coupling anti-MDDT antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing MDDT are passed over the immunoaffinity column, and the column is

washed under conditions that allow the preferential absorbance of MDDT (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/MDDT binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and MDDT is collected.

XVII. Identification of Molecules Which Interact with MDDT

MDDT, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled MDDT, washed, and any wells with labeled MDDT complex are assayed. Data obtained using different concentrations of MDDT are used to calculate values for the number, affinity, and association of MDDT with the candidate molecules.

Alternatively, molecules interacting with MDDT are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

MDDT may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of MDDT Activity

Phorbol ester binding activity of MDDT is measured using an assay based on the fluorescent phorbol ester sapinotoxin-D (SAPD). Binding of SAPD to MDDT is quantified by measuring the resonance energy transfer from MDDT tryptophans to the 2-(N-methylamino)benzoyl fluorophore of the phorbol ester, as described by Slater et al. (((1996) J. Biol. Chem. 271:4627-4631).

MDDT activity is associated with its ability to form protein-protein complexes and is measured by its ability to regulate growth characteristics of NIH3T3 mouse fibroblast cells. A cDNA encoding MDDT is subcloned into an appropriate eukaryotic expression vector. This vector is transfected into NIH3T3 cells using methods known in the art. Transfected cells are compared with non-transfected cells for the following quantifiable properties: growth in culture to high density, reduced attachment of cells to the substrate, altered cell morphology, and ability to induce tumors when injected into immunodeficient mice. The activity of MDDT is proportional to the extent of increased growth or frequency of altered cell morphology in NIH3T3 cells transfected with MDDT.

Alternatively, MDDT activity is measured by binding of MDDT to radiolabeled formin polypeptides containing the proline-rich region that specifically binds to SH3 containing proteins (Chan, D.C. et al. (1996) EMBO J. 15:1045-1054). Samples of MDDT are run on SDS-PAGE gels, and

transferred onto nitrocellulose by electroblotting. The blots are blocked for 1 hr at room temperature in TBST (137 mM NaCl, 2.7 mM KCl, 25 mM Tris (pH 8.0) and 0.1% Tween-20) containing non-fat dry milk. Blots are then incubated with TBST containing the radioactive formin polypeptide for 4 hrs to overnight. After washing the blots four times with TBST, the blots are exposed to autoradiographic film. Radioactivity is quantitated by cutting out the radioactive spots and counting them in a radioisotope counter. The amount of radioactivity recovered is proportional to the activity of MDDT in the assay.

Alternatively, MDDT protein kinase activity is measured by quantifying the phosphorylation of an appropriate substrate in the presence of gamma-labeled ^{32}P-ATP. MDDT is incubated with the substrate, ^{32}P-ATP, and an appropriate kinase buffer. The ^{32}P incorporated into the product is separated from free ^{32}P-ATP by electrophoresis, and the incorporated ^{32}P is quantified using a beta radioisotope counter. The amount of incorporated ^{32}P is proportional to the protein kinase activity of MDDT in the assay. A determination of the specific amino acid residue phosphorylated by protein kinase activity is made by phosphoamino acid analysis of the hydrolyzed protein.

Alternatively, an assay for MDDT protein phosphatase activity measures the hydrolysis of para-nitrophenyl phosphate (PNPP). MDDT is incubated together with PNPP in HEPES buffer pH 7.5, in the presence of 0.1% β-mercaptoethanol at 37°C for 60 min. The reaction is stopped by the addition of 6 ml of 10 N NaOH, and the increase in light absorbance of the reaction mixture at 410 nm resulting from the hydrolysis of PNPP is measured using a spectrophotometer. The increase in light absorbance is proportional to the activity of MDDT in the assay (Diamond, R.H. et al. (1994) Mol. Cell Biol. 14:3752-3762).

Alternatively, adenylyl cylcase activity of MDDT is demonstrated by the ability to convert ATP to cAMP (Mittal, C.K. (1986) Meth. Enzymol. 132:422-428). In this assay MDDT is incubated with the substrate [α-^{32}P]ATP, following which the excess substrate is separated from the product cyclic [^{32}P] AMP. MDDT activity is determined in 12 x 75 mm disposable culture tubes containing 5 μl of 0.6 M Tris-HCl, pH 7.5, 5 μl of 0.2 M $MgCl_2$, 5 μl of 150 mM creatine phosphate containing 3 units of creatine phosphokinase, 5 μl of 4.0 mM 1-methyl-3-isobutylxanthine, 5 μl of 20 mM cAMP, 5 μl 20 mM dithiothreitol, 5 μl of 10 mM ATP, 10 μl [α-^{32}P]ATP (2-4 x 10^6 cpm), and water in a total volume of 100 μl. The reaction mixture is prewarmed to 30°C. The reaction is initiated by adding MDDT to the prewarmed reaction mixture. After 10-15 minutes of incubation at 30°C, the reaction is terminated by adding 25 μl of 30% ice-cold trichloroacetic acid (TCA). Zero-time incubations and reactions incubated in the absence of MDDT are used as negative controls. Products are separated by ion exchange chromatography, and cyclic [^{32}P] AMP is quantified using a β-radioisotope counter. The MDDT activity is proportional to the amount of cyclic [^{32}P] AMP formed in the reaction.

An alternative assay measures MDDT-mediated G-protein signaling activity by monitoring the mobilization of Ca^{2+} as an indicator of the signal transduction pathway stimulation. (See, e.g., Grynkiewicz, G. et al. (1985) J. Biol. Chem. 260:3440; McColl, S. et al. (1993) J. Immunol. 150:4550-4555; and Aussel supra). The assay requires preloading neutrophils or T cells with a fluorescent dye such as FURA-2 or BCECF (Universal Imaging Corp, Westchester PA) whose emission characteristics are altered by Ca^{2+} binding. When the cells are exposed to one or more activating stimuli artificially (e.g., anti-CD3 antibody ligation of the T cell receptor) or physiologically (e.g., by allogeneic stimulation), Ca^{2+} flux takes place. This flux can be observed and quantified by assaying the cells in a fluorometer or fluorescent activated cell sorter. Measurements of Ca^{2+} flux are compared between cells in their normal state and those transfected with MDDT. Increased Ca^{2+} mobilization attributable to increased MDDT concentration is proportional to MDDT activity.

Alternatively, GTP-binding activity of MDDT is determined in an assay that measures the binding of MDDT to [α-^{32}P]-labeled GTP. Purified MDDT is first blotted onto filters and rinsed in a suitable buffer. The filters are then incubated in buffer containing radiolabeled [α-^{32}P]-GTP. The filters are washed in buffer to remove unbound GTP and counted in a radioisotope counter. Non-specific binding is determined in an assay that contains a 100-fold excess of unlabeled GTP. The amount of specific binding is proportional to the activity of MDDT.

Alternatively, GTPase activity of MDDT is determined in an assay that measures the conversion of [α-^{32}P]-GTP to [α-^{32}P]-GDP. MDDT is incubated with [α-^{32}P]-GTP in buffer for an appropriate period of time, and the reaction is terminated by heating or acid precipitation followed by centrifugation. An aliquot of the supernatant is subjected to polyacrylamide gel electrophoresis (PAGE) to separate GDP and GTP together with unlabeled standards. The GDP spot is cut out and counted in a radioisotope counter. The amount of radioactivity recovered in GDP is proportional to the GTPase activity of MDDT.

Alternatively, MDDT activity is measured by quantifying the amount of a non-hydrolyzable GTP analogue, GTPγS, bound over a 10 minute incubation period. Varying amounts of MDDT are incubated at 30 °C in 50 mM Tris buffer, pH 7.5, containing 1 mM dithiothreitol, 1 mM EDTA and 1 μM [^{35}S]GTPγS. Samples are passed through nitrocellulose filters and washed twice with a buffer consisting of 50 mM Tris-HCl, pH 7.8, 1 mM NaN_3, 10 mM $MgCl_2$, 1 mM EDTA, 0.5 mM dithiothreitol, 0.01 mM PMSF, and 200 mM NaCl. The filter-bound counts are measured by liquid scintillation to quantify the amount of bound [^{35}S]GTPγS. MDDT activity may also be measured as the amount of GTP hydrolysed over a 10 minute incubation period at 37 °C. MDDT is incubated in 50mM Tris-HCl buffer, pH 7.8, containing 1mM dithiothreitol, 2mM EDTA, 10μM [α-^{32}P]GTP, and 1

μM H-rab protein. GTPase activity is initiated by adding $MgCl_2$ to a final concentration of 10 mM. Samples are removed at various time points, mixed with an equal volume of ice-cold 0.5mM EDTA, and frozen. Aliquots are spotted onto polyethyleneimine-cellulose thin layer chromatography plates, which are developed in 1M LiCl, dried, and autoradiographed. The signal detected is proportional to MDDT activity.

Alternatively, MDDT activity may be demonstrated as the ability to interact with its associated low molecular weight (LMW) GTPase in an in vitro binding assay. The candidate LMW GTPases are expressed as fusion proteins with glutathione S-transferase (GST), and purified by affinity chromatography on glutathione-Sepharose. The LMW GTPases are loaded with GDP by incubating 20 mM Tris buffer, pH 8.0, containing 100 mM NaCl, 2 mM EDTA, 5 mM $MgCl_2$, 0.2 mM DTT, 100 µM AMP-PNP and 10 µM GDP at 30°C for 20 minutes. MDDT is expressed as a FLAG fusion protein in a baculovirus system. Extracts of these baculovirus cells containing MDDT-FLAG fusion proteins are precleared with GST beads, then incubated with GST-GTPase fusion proteins. The complexes formed are precipitated by glutathione-Sepharose and separated by SDS-polyacrylamide gel electrophoresis. The separated proteins are blotted onto nitrocellulose membranes and probed with commercially available anti-FLAG antibodies. MDDT activity is proportional to the amount of MDDT-FLAG fusion protein detected in the complex.

Another alternative assay to detect MDDT activity is the use of a yeast two-hybrid system (Zalcman, G. et al. (1996) J. Biol. Chem. 271:30366-30374). Specifically, a plasmid such as pGAD1318 which may contain the coding region of MDDT can be used to transform reporter L40 yeast cells which contain the reporter genes *LacZ* and *HIS3* downstream from the binding sequences for LexA. These yeast cells have been previously transformed with a pLexA-Rab6-GDP (mouse) plasmid or with a plasmid which contains pLexA-lamin C. The pLEXA-lamin C cells serve as a negative control. The transformed cells are plated on a histidine-free medium and incubated at 30°C for 3 days. His^+ colonies are subsequently patched on selective plates and assayed for β-galactosidase activity by a filter assay. MDDT binding with Rab6-GDP is indicated by positive $His^+/lacZ^+$ activity for the cells transformed with the plasmid containing the mouse Rab6-GDP and negative $His^+/lacZ^+$ activity for those transformed with the plasmid containing lamin C.

Alternatively, MDDT activity is measured by binding of MDDT to a substrate which recognizes WD-40 repeats, such as ElonginB, by coimmunoprecipitation (Kamura, T. et al. (1998) Genes Dev. 12:3872-3881). Briefly, epitope tagged substrate and MDDT are mixed and immunoprecipitated with commercial antibody against the substrate tag. The reaction solution is run on SDS-PAGE and the presence of MDDT visualized using an antibody to the MDDT tag. Substrate binding is proportional to MDDT activity.

Alternatively, MDDT activity is measured by its inclusion in coated vesicles. MDDT can be expressed by transforming a mammalian cell line such as COS7, HeLa, or CHO with a eukaryotic expression vector encoding MDDT. Eukaryotic expression vectors are commercially available, and the techniques to introduce them into cells are well known to those skilled in the art. A small amount of a second plasmid, which expresses any one of a number of marker genes, such as β-galactosidase, is co-transformed into the cells in order to allow rapid identification of those cells which have taken up and expressed the foreign DNA. The cells are incubated for 48-72 hours after transformation under conditions appropriate for the cell line to allow expression and accumulation of MDDT and β-galactosidase.

In the alternative, MDDT activity is measured by its ability to alter vesicle trafficking pathways. Vesicle trafficking in cells transformed with MDDT is examined using fluorescence microscopy. Antibodies specific for vesicle coat proteins or typical vesicle trafficking substrates such as transferrin or the mannose-6-phosphate receptor are commercially available. Various cellular components such as ER, Golgi bodies, peroxisomes, endosomes, lysosomes, and the plasmalemma are examined. Alterations in the numbers and locations of vesicles in cells transformed with MDDT as compared to control cells are characteristic of MDDT activity. Transformed cells are collected and cell lysates are assayed for vesicle formation. A non-hydrolyzable form of GTP, GTPγS, and an ATP regenerating system are added to the lysate and the mixture is incubated at 37 °C for 10 minutes. Under these conditions, over 90% of the vesicles remain coated (Orci, L. et al. (1989) Cell 56:357-368). Transport vesicles are salt-released from the Golgi membranes, loaded under a sucrose gradient, centrifuged, and fractions are collected and analyzed by SDS-PAGE. Co-localization of MDDT with clathrin or COP coatamer is indicative of MDDT activity in vesicle formation. The contribution of MDDT in vesicle formation can be confirmed by incubating lysates with antibodies specific for MDDT prior to GTPγS addition. The antibody will bind to MDDT and interfere with its activity, thus preventing vesicle formation.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
4973222	1	4973222CD1	40	4973222CB1	
55009060	2	55009060CD1	41	55009060CB1	6607188CA2
1985092	3	1985092CD1	42	1985092CB1	3269039CA2
1553593	4	1553593CD1	43	1553593CB1	90089320CA2, 90089412CA2
1954122	5	1954122CD1	44	1954122CB1	1534622CA2, 90104942CA2
3159276	6	3159276CD1	45	3159276CB1	
140052	7	140052CD1	46	140052CB1	
5158048	8	5158048CD1	47	5158048CB1	
3127541	9	3127541CD1	48	3127541CB1	
8224777	10	8224777CD1	49	8224777CB1	
587394	11	587394CD1	50	587394CB1	
1402405	12	1402405CD1	51	1402405CB1	90088561CA2, 90088677CA2
1798468	13	1798468CD1	52	1798468CB1	90089911CA2, 90089943CA2, 90090003CA2, 90090035CA2
3189084	14	3189084CD1	53	3189084CB1	
5580384	15	5580384CD1	54	5580384CB1	90096580CA2, 90096656CA2
5158619	16	5158619CD1	55	5158619CB1	5279344CA2, 5301408CA2, 90005696CA2

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
2792745	17	2792745CD1	56	2792745CB1	90086574CA2, 90086590CA2, 90086674CA2, 90086749CA2, 90086757CA2, 90086765CA2, 90086781CA2, 90086789CA2, 90086849CA2, 90086857CA2, 90086865CA2, 90086873CA2, 90086881CA2, 90086889CA2
2827678	18	2827678CD1	57	2827678CB1	
790257	19	790257CD1	58	790257CB1	
2617345	20	2617345CD1	59	2617345CB1	2617345CA2
3254666	21	3254666CD1	60	3254666CB1	2013975CA2, 6549832CA2
4159378	22	4159378CD1	61	4159378CB1	90090796CA2, 90090896CA2
4317538	23	4317538CD1	62	4317538CB1	
1881010	24	1881010CD1	63	1881010CB1	3218687CA2
1593038	25	1593038CD1	64	1593038CB1	
7494930	26	7494930CD1	65	7494930CB1	
7497349	27	7497349CD1	66	7497349CB1	
5510805	28	5510805CD1	67	5510805CB1	5510805CA2
1577482	29	1577482CD1	68	1577482CB1	
1805054	30	1805054CD1	69	1805054CB1	

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
7492708	31	7492708CD1	70	7492708CB1	
7490847	32	7490847CD1	71	7490847CB1	
7493059	33	7493059CD1	72	7493059CB1	
2321130	34	2321130CD1	73	2321130CB1	
2008365	35	2008365CD1	74	2008365CB1	
3580778	36	3580778CD1	75	3580778CB1	8012007CA2
7948785	37	7948785CD1	76	7948785CB1	
7494415	38	7494415CD1	77	7494415CB1	
2234223	39	2234223CD1	78	2234223CB1	2234223CA2

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	4973222CD1	g29495	1.8E-97	[Homo sapiens] B cell stimulatory factor-2 (BSF-2) Yasukawa, K. et al. (1987) Structure and expression of human B cell stimulatory factor-2 (BSF-2/IL-6) gene. EMBO J. 6:2939-2945
2	55009060CD1	g505033	1.2E-197	[Homo sapiens] mitogen inducible gene mig-2 Wick, M. et al. (1994) Identification of serum-inducible genes: Different patterns of gene regulation during G0-->S and G1-->S progression. J. Cell. Sci. 107 (Pt 1):227-239
5	1954122CD1	g4539599	1.9E-13	[Schizosaccharomyces pombe] WD repeat protein
8	5158048CD1	g15420726	0.0	[Mus musculus] melanophilin Matesic,L.E., et al. (2001) Mutations in Mlph, encoding a member of the Rab effector family, cause the melanosome transport defects observed in leaden mice. Proc. Natl. Acad. Sci. U.S.A. 98:10238-10243
9	3127541CD1	g2624972	9.1E-65	[Mus musculus] proline-rich protein 48 Ermekova, K.S. et al. (1997) The WW domain of neural protein FE65 interacts with proline-rich motifs in mena, the mammalian homolog of Drosophila enabled. J. Biol. Chem. 272:32869-32877
10	8224777CD1	g7211438	1.6E-157	[Homo sapiens] golgin-67 Eystathioy, T., et al. (2000) Human autoantibodies to a novel Golgi protein golgin-67: high similarity with golgin-95/gm 130 autoantigen. J. Autoimmun. 14, 179-187
14	3189084CD1	g817954	0.0	[Mus musculus] DMR-N9 Jansen, G., et al. (1995) Structural organization and developmental expression pattern of the mouse WD-repeat gene DMR-N9 immediately upstream of the myotonic dystrophy locus. Hum. Mol. Genet. 4, 843-852.
15	5580384CD1	g11094230	1.7E-50	[Oryctolagus cuniculus] RPBP1 (rabbit placenta basic protein 1)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
16	5158619CD1	g1799568	2.2E-32	[Homo sapiens] stac Suzuki, H., et al. (1996) Stac, a novel neuron-specific protein with cysteine-rich and SH3 Domains. Biochem. Biophys. Res. Commun. 229, 902-909
17	2792745CD1	g1872200	6.5E-21	[Homo sapiens] alternatively spliced product using exon 13A
18	2827678CD1	g1872200	1.3E-27	[Homo sapiens] alternatively spliced product using exon 13A
23	4317538CD1	g17907791	0.0	[Homo sapiens] TGF-beta induced apotosis protein 2
24	1881010CD1	g13936285	0.0	[Mus musculus] TRH4
25	1593038CD1	g13936275	0.0	[Mus musculus] RANBP20
26	7494930CD1	g339777	0.0	[Homo sapiens] ORF2 contains a reverse transcriptase domain. Santos, F.R. et al (2000) Hum. Mol. Genet. 9 (3), 421-430
27	7497349CD1	g20135652	3.0E-73	[Homo sapiens] BRAF35/HDAC2 complex 80 kDa protein Hakimi,M.-A. (2002) A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific gene. Proc. Natl. Acad. Sci. U.S.A. In press
30	1805054CD1	g11993616	0.0	[Homo sapiens] MTO1 protein
31	7492708CD1	g483915	1.5E-159	[Homo sapiens] ORF1, encodes a 40 kDa product Holmes, S.E. et al (1994) Nature Genet. 7:143-148
32	7490847CD1	g10441006	2.6E-17	[Xenopus laevis] 4g2
33	7493059CD1	g339771	0.0	[Homo sapiens] ORF2 contains a reverse transcriptase domain; ORF2
38	7494415CD1	g2072957	8.6E-134	[Homo sapiens] p40 Sassaman,D.M. et al. (1997) Many human L1 elements are capable of retrotransposition. Nature Genet. 16:37-43
39	2234223CD1	g15420726	4.7E-115	[Mus musculus] melanophilin Matesic,L.E., et al. (2001) Mutations in Mlph, encoding a member of the Rab effector family, cause the melanosome transport defects observed in leaden mice. Proc. Natl. Acad. Sci. U.S.A. 98:10238-10243

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	4973222CD1	200	S38 S68 S123 S134 S185 S193 T36 T59 T153 T178	N61 N160	signal_cleavage: M1-P15	SPSCAN
					Interleukin-6/G-CSF/MGF family: I45-R198	HMMER_PFAM
					Interleukin-6/G-CSF/MGF proteins BL00254: E75-N119, Q168-R198	BLIMPS_BLOCKS
					Interleukin-6/G-CSF/MGF family signature interleukin_6.prf: S69-S134	PROFILESCAN
					Interleukin-6/G-CSF/MGF family signature PR00433: Q44-T59, C60-M83, C89-L114, I182-R198	BLIMPS_PRINTS
					Interleukin-6 signature PR00434: Q44-C60, C66-G88, C89-F110, L181-R198	BLIMPS_PRINTS
					FACTOR GROWTH GLYCOPROTEIN CYTOKINE PRECURSOR SIGNAL INTERLEUKIN6 IL6 GRANULOCYTE COLONY STIMULATING PD004356: I45-R198	BLAST_PRODOM
					INTERLEUKIN-6/G-CSF/MGF DM02670|P46650|1-211: M1-M200 DM02670|P41323|1-206: M1-M200 DM02670|P41683|1-207: M1-M200 DM02670|P26892|1-207: M1-R198	BLAST_DOMO
					Interleukin-6/G-CSF/MGF signature: C89-L114	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2	55009060CD1	663	S14 S132 S218 S220 S232 S255 S328 S386 S531 S545 S561 S623 S638 T33 T287 T348 T365 T456 T517 T587 Y162 Y441 Y462	N410	PH domain: P350-K453	HMMER_PFAM
					Transmembrane Domain: A109-H135 N-terminus is non-cytosolic	TMAP
					C47E8.7 UNC-112 cell matrix adhesion structure protein PD147334: I12-R143, R215-G659	BLAST_PRODOM
3	1985092CD1	219	S2 S60		Transmembrane Domain: P31-L56 E65-D93 G109-A137 P140-Y168 N-terminus is non-cytosolic	TMAP
4	1553593CD1	318	S28 S35 S65 S113 S115 S141 S164 S239 T218			
5	1954122CD1	387	S51 S225 S362 S380 T31 T93 T134 T157 T168 T236 T295 T323 T338 T358	N49 N154 N304 N378	WD domain, G-beta repeat: C16-D56, V62-D98, L162-D199, L209-D245, V313-K349	HMMER_PFAM
					Trp-Asp (WD-40) repeats signature: S181-S220	PROFILESCAN
					G-protein beta WD-40 repeat signature PR00320: V85-A99, V186-I200	BLIMPS_PRINTS
					Protein Repeat WD TrpAsp Repeats Containing Chromosome Nuclear Factor I PD000061: V76-A99, D140-I200	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5 (cont.)					Trp-Asp (WD) repeat protein BL00678: S87-W97	BLIMPS_BLOCKS
6	3159276CD1	577	S27 S41 S73 S74 S237 S259 S447 S449 S478 S493 S509 S527 S531 S545 S562 S569 T118 T379 Y112 Y572	N223 N365		
7	140052CD1	224	S24 S121 S160 S183 S188 T105 T139 T168 T217			
8	5158048CD1	600	S47 S48 S77 S134 S184 S191 S216 S226 S231 S233 S256 S282 S314 S318 S329 S349 S397 S402 S403 S444 S463 S484 S503 S510 S552 S554 T11 T88 T165 T313 T356 T376 T401 T438 T457 T458 T472 T499 T576	N60 N395 N455	TROPOMYOSIN DM00077|P42638|29-206: T376-K501	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	3127541CD1	1250	S5 S17 S44 S66 S108 S140 S152 S166 S174 S197 S217 S281 S313 S412 S522 S533 S542 S590 S696 S790 S961 S980 S985 S996 S1047 S1054 S1060 S1069 S1098 S1115 S1130 S1154 S1185 S1189 T118 T127 T132 T195 T243 T292 T317 T373 T395 T585 T960 T1134 T1196 T1228 Y524	N64 N206 N339 N549 N552 N675 N978 N1031 N1128 N1170	PH domain: P397-Y505	HMMER_PFAM
					PROTEIN REPEAT SIGNAL PRECURSOR PRION GLYCOPROTEIN NUCLEAR GPI ANCHOR BRAIN MAJOR PD001091: Q800-P1021	BLAST_PRODOM
					PROTEIN REPEAT MICROTUBULE ASSOCIATED MICROTUBULES PHOSPHORYLATION BASSOON ALTERNATIVE SPLICING LARGE PROLINE-RICH PD005493: A742-A1022	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9 (cont.)					RECEPTOR PROTEIN BINDING GROWTH FACTOR INSULIN GRB10 ADAPTER GRB7 SH2 PD007754: I399-I543	BLAST_PRODOM
					BAT2; Large Proline-Rich Protein DM05517|P48634|1-1860: P607-P1099 DM05517|S37671|1-1870: P607-P1099	BLAST_DOMO
					Verprolin DM08461|P37370|203-451: S736-P966	BLAST_DOMO
10	8224777CD1	621	S73 S110 S122 S175 S233 S324 S392 S430 S447 S542 T16 T30 T119 T210 T226 T285	N401	signal_cleavage: M1-A65	SPSCAN
					GOLGI STACK COILED COIL GOLGIN95 CISGOLGI MATRIX PROTEIN GM130 SIMILAR PD033411: N529-Q594	BLAST_PRODOM
					PROTEIN COILED COIL CHAIN MYOSIN REPEAT HEAVY ATPBINDING FILAMENT HEPTAD PD000002: I160-L418	BLAST_PRODOM
					GOLGIN95 GOLGI STACK COILED COIL PD173178: E249-M311	BLAST_PRODOM
					TRICHOHYALIN DM03839|P37709|632-1103: Q100-E413	BLAST_DOMO
					CAP-GLY DOMAIN DM03881|P28023|177-1231: P64-Q410	BLAST_DOMO
11	587394CD1	114	S22 S26 S37 S49 S73 S86	N83	signal_cleavage: M1-C25	SPSCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11 (cont.)					Transmembrane domain: Q78-V94 N-terminus is non-cytosolic	TMAP
12	1402405CD1	527	S187 S225 S441 T50 T55 T180 T287 T493		Transmembrane domain: A397-L414 N-terminus is non-cytosolic	TMAP
					NEGATIVE FACTOR F-PROTEIN PD00444: P113-L151, G192-V210, D224-A269	BLIMPS_PRODOM
13	1798468CD1	316	S82 S91 S118 T20 T23 T30 T56 T130 T185 T286 T300 T305 T310 Y117	N52	PX domain: K72-P187	HMMER_PFAM
14	3189084CD1	659	S106 S130 S339 S382 S460 S544 S556 S641 T151 T186 T328 T365 T366 T413 T512 T621 Y18 Y145	N227 N533	WD domain, G-beta repeat: P261-H297, L306-S339, V345-D378, N191-N227	HMMER_PFAM
					Trp-Asp (WD) repeat protein signature BL00678: T328-W338	BLIMPS_BLOCKS
					PROTEIN DMRN9 C08B6.7 PD040879: L403-K640 PD022683: S106-G313	BLAST_PRODOM
					BETA-TRANSDUCIN FAMILY TRP-ASP REPEATS DM00005|Q08274|230-294: S234-H297 DM00005|Q08274|295-340: F298-G343	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
15	5580384CD1	446	S117 S128 S130 S139 S220 S240 S286 S318 S334 S375 S378 S414 T44 T80 T94 T166 T218 T259	N70 N284	G-patch domain: G11-D55	HMMER_PFAM
					PROTEIN REPEAT TROPOMYOSIN COILED COIL ALTERNATIVE SPLICING SIGNAL PRECURSOR CHAIN PD000023: Q194-D399	BLAST_PRODOM
					TRICHOHYALIN DM03839|Q07283|91-443: K170-K444	BLAST_DOMO
16	5158619CD1	364	S9 S34 S133 S213 S280 T2 T180 T236 T275 T327 T349 Y344		Phorbol esters/diacylglycerol binding domain: H90-C140	HMMER_PFAM
					SH3 domain: H250-V304	HMMER_PFAM
					Phorbol esters / diacylglycerol binding domain BL00479: H90-L112, F116-C131	BLIMPS_BLOCKS
					Src homology 3 (SH3) domain BL50002: A254-E272, G290-R303	BLIMPS_BLOCKS
					Phorbol esters / diacylglycerol binding domain: F102-Q160	PROFILESCAN
					SH3 domain signature PR00452: N281-G290, H250-A260, D264-D279	BLIMPS_PRINTS
					Phorbol esters / diacylglycerol binding domain: H90-C140	MOTIFS
17	2792745CD1	91	S36 S46 T71		Transmembrane domain: N50-L67 N-terminus is cytosolic.	TMAP

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17 (cont.)					PROTEIN ALU SUBFAMILY RNA EDITING PROTOONCOGENE REPEAT: PD005171: S5-N50	BLAST_PRODOM
18	2827678CD1	116	S49 S55 T90		signal_cleavage: M1-A28	SPSCAN
					Signal Peptides: M1-A26, M1-A28	HMMER
					Transmembrane domains: F8-G36, S56-I81 N-terminus is non-cytosolic.	TMAP
					PROTEIN ALU SUBFAMILY RNA EDITING PROTOONCOGENE REPEAT: PD005171: V25-N68	BLAST_PRODOM
					PROTEIN PROTO-ONCOGENE NUCLEAR UBIQUITOUS TPR MOTIF Y ISOFORM: PD015557: F70-A111	BLAST_PRODOM
19	790257CD1	684	S42 S189 S199 S225 S438 S467 S543 S571 S648 S659 T262 T319 T339 T416 T422 T459 T462 T520 T635 T675	N148 N527 N565	Transmembrane domains: T150-P178, S467-E484, V615-K643 N-terminus is non-cytosolic.	TMAP
					Sec1 family PF00995: K398-M444, P624-V646, T653-R677	BLIMPS_PFAM
20	2617345CD1	344	S204 S228 S281 T92 T174 T202	N152	WD domain, G-beta repeat: E223-A258, V134-D170, N2-N37, P177-R211	HMMER_PFAM
					PROTEIN INTERGENIC REGION PD013323: N7-G160	BLAST_PRODOM
21	3254666CD1	95	T14 Y89			

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
22	4159378CD1	410	S142 S194 S248 S273 S320 S407 T5 T53 T84 T157 T172 T246 T338 T369 T391	N141	Transmembrane domain: K7-K29 N-terminus is non-cytosolic.	TMAP
					PROTEIN COSMID CHROMOSOME III: PD013551: E198-L408	BLAST_PRODOM
					Leucine zipper pattern: L82-L103, L364-L385	MOTIFS
23	4317538CD1	616	S51 S55 S60 S126 S131 S163 S188 S239 S343 S363 S398 S400 S486 S536 S538 S594 T107 T136 T183 T197 T210 T430 T490 Y519	N171 N456 N462 N484 N493 N540	RAT MITOCHONDRIAL CAPSULE SELENO-PROTEIN PD144344: D185-V434, S46-L137	BLAST_PRODOM
24	1881010CD1	392	S18 S28 S345 S350 S354 S355 S356 S367 T296	N26 N294	Homeobox domain: L92-P135	HMMER_PFAM
					Transmembrane domains: R45-K70, C142-S165, Y190-K210, I216-M236, A263-W291, A308-L336 N-terminus is non-cytosolic.	TMAP
					PROTEIN TRANSMEMBRANE LONGEVITY ASSURANCE FACTOR S CEREVISIAE PD006418: H128-K366, W175-K366	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
25	1593038CD1	1125	S7 S54 S227 S301 S358 S366 S416 S553 S666 S790 S803 S813 T23 T55 T66 T152 T153 T193 T210 T355 T429 T477 T585 T656 T687 T1071 T1122 Y1110	N293 N540 N798 N1115	Transmembrane domains: E171-H198, A249-F277, S559-A583, H679-I696, K1015-H1039, A1049-T1071 N-terminus is cytosolic.	TMAP
26	7494930CD1	1273	S79 S156 S202 S335 S454 S455 S469 S506 S1043 S1079 S1213 T2 T47 T151 T216 T226 T243 T249 T352 T382 T391 T409 T466 T482 T523 T772 T794 T836 T971 T974 T982 T996 T1019 T1032 T1063 T1190 Y97	N108 N134 N169 N245 N258 N277 N360 N387 N449 N722 N898	AP endonuclease family: I8-R238	HMMER_PFAM
					Reverse transcriptase (RNA-dependent DNA polymerase): G501-L771	HMMER_PFAM
					Transmembrane domain: L651-R678, C802-I820 N-terminus is non-cytosolic	TMAP
					AP endonucleases family BL00726: P36-L46, F217-I239	BLIMPS_BLOCKS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
26 (cont.)					DNA RNADIRECTED POLYMERASE PUTATIVE P150 TRANSCRIPTASE REVERSE PROTEIN L1 SEQUENCE PD002894: V153-Q353 PD002747: M1074-T1214 PD003002: T772-W886 PD002970: I905-G987	BLAST_PRODOM
					TRANSCRIPTASE; REVERSE; II; ORF2; DM01354|I38588|559-974: I557-K973 DM01354|S23650|1-411: V562-K973 DM01354|P08547|558-973: I557-K973 DM01354|JU0033|48-463: I557-K973	BLAST_DOMO
27	7497349CD1	327	S13 S40 S94 S103 S173 S273 T36 T58 T104 T115 T132 T146 T217 T220		PHD-finger: H150-K195	HMMER_PFAM
					C3A-anaphylatoxin receptor signature PR01060: P72-T86, D247-Q272, Q292-K310	BLIMPS_PRINTS
					Pterin 4 alpha carbinolamine dehydratase PF01329: L243-Q255	BLIMPS_PFAM
					HOMEODOMAIN; PATHOGENESIS; YMR075W; DM02014|Q09819|49-172: K142-A196	BLAST_DOMO
					Leucine zipper pattern: L229-L250	MOTIFS
28	5510805CD1	79	S3	N42	Transmembrane domain: M47-L75 N-terminus is non-cytosolic	TMAP

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
29	1577482CD1	270	S31 S37 S131 S172 S237 T20 T74 T127	N50		
30	1805054CD1	692	S429 S433 S441 S468 S475 S509 S510 S525 S669 S674 S675 T66 T149 T161 T168 T308 T412 T467 T486 T539 T573 T618 T664 Y134 Y553 Y588	N249 N269 N409 N427 N672	signal_cleavage: M1-A51	SPSCAN
					Glucose inhibited division protein A: F37-Q665	HMMER_PFAM
					Pyridine nucleotide-disulphide oxidoreductase: V39-L65	HMMER_PFAM
					Glucose inhibited division protein A family proteins BL01280: L454-R500, K612-L658, D38-N78, D97-L148, G175-L225, R234-F250, Q279-G311, G311-G338, Y382-A428	BLIMPS_BLOCKS
					Pyridine nucleotide disulphide reductase class-II signature PR00469: D38-S60, E187-T195	BLIMPS_PRINTS
					FAD-dependent glycerol-3-phosphate dehydrogenase family signature PR01001: F37-T49	BLIMPS_PRINTS
					PROTEIN GLUCOSE INHIBITED DIVISION A GID FAD-DEPENDENT OXIDOREDUCTASE PI079 HAP2ADE5 PD003738: F37-K662	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
30 (cont.)					GIDA PROTEIN DM01921|P53070|201-678: I201-S669 DM01921|P25756|163-630: I201-K662 DM01921|P17112|163-628: I201-A670 DM01921|I64078|164-630: I203-L658	BLAST_DOMO
					Glucose inhibited division protein A family signature 1: G311-F325	MOTIFS
					Glucose inhibited division protein A family signature 2: A405-A428	MOTIFS
31	7492708CD1	338	S6 S12 S16 S33 S106 S109 S119 S145 S166 S208 S209 S290 T35 T81 T83 T172 T177 T213 T222 T241 T246 T298 T308	N51 N170	L1 ELEMENT P40 PUTATIVE P150 GENES COMPLETE CDS A L1.8 PD005182: M230-L338 PD005183: R49-R138	BLAST_PRODOM
					ELEMENT L1 P40 PUTATIVE P150 GENES COMPLETE CDS A L1.8 PD004752: M1-R48	BLAST_PRODOM
					L1 ELEMENT P40 PUTATIVE P150 GENES COMPLETE CDS PROTEIN A PD003272: I139-K229	BLAST_PRODOM
					ORF1; KDA; RNA; PRODUCT; DM01447|A34087|53-236: D151-H335 DM01447|I38587|151-334: D151-H335 DM01447|A28096|151-334: D151-H335 DM01447|S21345|179-350: N176-N334	BLAST_DOMO
					Leucine zipper pattern: L93-L114	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
32	7490847CD1	1027	S10 S21 S59 S218 S259 S269 S329 S347 S359 S523 S590 S595 S686 S727 S794 S804 S828 S925 S967 T145 T165 T280 T308 T351 T401 T450 T484 T538 T544 T648 T663 T811 T874 Y74	N237 N474 N570 N725 N1018		
33	7493059CD1	1275	S79 S151 S156 S202 S312 S335 S457 S509 S763 S862 S1045 S1081 S1096 S1109 T47 T51 T238 T243 T249 T276 T290 T352 T393 T411 T455 T468 T500 T525 T526 T796 T838 T844 T973 T976 T998 T1021 T1026 T1034 T1065 T1215 Y626	N133 N245 N361 N545 N588 N724 N900	AP endonuclease family: I8-T238	HMMER_PFAM
					Reverse transcriptase (RNA-dependent): G504-L773	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
33 (cont.)					Transmembrane domains: R658-R680, K798-N826 N-terminus is cytosolic.	TMAP
					DNA RNA-DIRECTED POLYMERASE PUTATIVE P150 TRANSCRIPTASE REVERSE PROTEIN L1 SEQUENCE: PD002894: L153-Q353 PD002747: M1076-V1217 PD003002: T774-W888 PD003182: I41-T152	BLAST_PRODOM
					TRANSCRIPTASE; REVERSE; II; ORF2: DM01354|P08547|558-973: I560-K975 DM01354|I38588|559-974: I560-K975 DM01354|S23650|1-411: V565-K975 DM01354|JU0033|48-463: I560-K975	BLAST_DOMO
34	2321130CD1	635	S97 S115 S146 S196 S202 S325 S377 S392 S441 S442 T287 T299 T361 T395 T567	N25 N324 N333 N376 N440	signal_cleavage: M1-A16	SPSCAN
					Signal Peptides: M1-A26, M1-A28	HMMER
					Leucine Rich Repeat: S146-P169, S97-V120, G73-E96, R49-T72, Q194-G217, A170-G193, N121-L144	HMMER_PFAM
					Leucine rich repeat C-terminal domain: N234-E279	HMMER_PFAM
					Fibronectin type III domain: E404-T487	HMMER_PFAM
					Immunoglobulin domain: G295-A353	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
34 (cont.)					Transmembrane domain: A511-V539 N-terminus is cytosolic.	TMAP
					Leucine_Zipper: L19-L40	MOTIFS
35	2008365CD1	170	S84 S113			
36	3580778CD1	388	S45 S63 S115 S221 S231 S234 S277 T42 T148 T286 T288 T351 Y235 Y252	N111 N219	DnaJ domain: E51-Q114	HMMER_PFAM
37	7948785CD1	347	S12 S126 S223 S262 S270 T117 T122 T138 T229 T255 T290	N179 N196	Transmembrane domain: P13-K29 N-terminus is non-cytosolic.	TMAP
38	7494415CD1	338	S12 S109 S119 S145 S166 S208 S209 T33 T81 T96 T172 T177 T213 T222 T241 T298 T308 Y296	N16 N170	L1 ELEMENT P40 PUTATIVE P150 GENES COMPLETE: PD005182: M230-M338 PD005183: R48-R138 PD003272: I139-K229	BLAST_PRODOM
					Leucine_Zipper: L69-L90	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
39	2234223CD1	520	S47 S48 S77 S134 S184 S191 S216 S226 S231 S233 S256 S282 S314 S318 S329 S369 S374 S375 S380 S404 S423 S430 S472 S474 T11 T88 T165 T313 T348 T373 T419 T496	N60 N367		

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
40/4973222CB1/733	1-265, 1-266, 170-334, 191-442, 216-733, 266-518, 298-561
41/55009060CB1/2502	1-254, 1-406, 1-471, 1-660, 1-662, 1-691, 2-368, 6-304, 9-470, 13-249, 22-359, 24-511, 26-421, 27-285, 27-532, 179-488, 265-711, 272-683, 334-551, 334-834, 370-555, 399-681, 408-651, 468-795, 519-1076, 557-789, 574-873, 582-798, 623-891, 748-918, 748-1309, 753-1050, 757-989, 764-1008, 770-991, 898-1163, 904-1145, 1025-1195, 1026-1303, 1026-1519, 1034-1262, 1034-1524, 1081-1289, 1081-1542, 1193-1443, 1290-1585, 1316-1567, 1316-1716, 1319-1575, 1323-1495, 1366-1632, 1378-1649, 1380-1633, 1416-1671, 1440-1684, 1444-1692, 1484-1798, 1511-1871, 1517-1801, 1552-1817, 1677-2365, 1687-1935, 1688-1924, 1744-2380, 1770-1981, 1775-1990, 1776-2000, 1789-2039, 1809-2098, 1809-2475, 1878-2465, 1884-2172, 1899-2449, 1910-2120, 1910-2150, 1915-2467, 1932-2479, 1967-2485, 1997-2190, 1997-2198, 2213-2487, 2225-2473, 2241-2407, 2248-2502
42/1985092CB1/1591	1-148, 1-151, 1-166, 1-189, 1-441, 81-297, 81-603, 107-381, 115-347, 115-375, 148-244, 148-401, 192-382, 192-483, 196-443, 436-528, 446-670, 446-971, 505-532, 509-728, 509-729, 509-1128, 518-1121, 519-759, 527-750, 547-787, 598-849, 668-822, 754-1130, 758-969, 763-1020, 763-1237, 763-1286, 778-1068, 809-1058, 817-1023, 819-1021, 819-1094, 819-1331, 819-1357, 878-1154, 887-1556, 914-1151, 950-1556, 952-1326, 964-1562, 970-1555, 989-1556, 1027-1345, 1038-1556, 1047-1302, 1047-1536, 1096-1303, 1096-1350, 1124-1568, 1146-1545, 1148-1582, 1154-1518, 1166-1513, 1167-1549, 1174-1580, 1181-1395, 1181-1570, 1188-1439, 1194-1539, 1197-1591, 1220-1580, 1231-1479, 1234-1506, 1241-1500, 1254-1555, 1265-1567, 1306-1580, 1309-1580, 1323-1441, 1323-1519, 1323-1556, 1323-1591, 1369-1591, 1389-1570, 1389-1591, 1419-1568, 1421-1551
43/1553593CB1/1210	1-251, 1-271, 1-305, 1-308, 30-308, 34-176, 42-142, 46-276, 56-278, 57-249, 57-380, 66-285, 70-313, 70-315, 77-368, 77-675, 79-356, 85-181, 90-190, 104-315, 252-918, 368-575, 376-662, 380-678, 395-654, 427-927, 461-1210, 579-751, 701-1210, 797-1122, 810-894, 817-1049, 836-1083, 836-1210, 857-1069, 881-1068
44/1954122CB1/3112	1-223, 1-407, 20-261, 21-316, 52-494, 60-397, 149-400, 149-673, 307-801, 308-845, 309-583, 309-855, 388-657, 514-1052, 616-900, 641-886, 657-885, 715-1173, 852-1095, 961-1204, 978-1294, 1048-1696, 1108-1483, 1118-1590, 1139-1702, 1179-1590, 1191-1610, 1193-1428, 1233-1699, 1248-1538, 1298-1729, 1376-1699, 1381-1661, 1381-1669, 1497-1972, 1541-1990, 1548-1871, 1548-1929, 1569-1973, 1600-1865, 1607-2036, 1738-2030, 1850-2088, 1850-2106, 1917-2475, 1929-2200, 1981-2273, 2006-2581, 2008-2350, 2057-2386, 2173-2428, 2216-2440, 2294-2548, 2302-2865, 2356-2566, 2389-3112, 2446-2697, 2518-2793, 2549-2777, 2566-2857, 2608-2882, 2611-2863, 2612-2854

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
45/3159276CB1/2398	1-558, 1-2345, 30-578, 219-645, 329-2363, 345-761, 345-785, 354-772, 364-992, 381-768, 384-938, 423-677, 423-878, 427-680, 427-719, 455-701, 494-799, 562-701, 565-1052, 573-806, 606-864, 660-1387, 683-1193, 702-1307, 711-1307, 724-1005, 732-1400, 740-1365, 750-1418, 777-1511, 786-1454, 816-1444, 841-1429, 847-1515, 873-1307, 873-1399, 892-1545, 893-1515, 896-1463, 896-1521, 902-1540, 902-1560, 912-1530, 921-1465, 928-1398, 942-1429, 954-1210, 954-1407, 957-1506, 957-1509, 959-1502, 962-1522, 963-1604, 968-1467, 971-1486, 976-1493, 981-1388, 991-1024, 995-1604, 1007-1494, 1010-1251, 1011-1250, 1014-1292, 1024-1548, 1077-1614, 1088-1646, 1093-1307, 1106-1608, 1122-1789, 1132-1735, 1133-1409, 1191-1795, 1198-1486, 1210-1458, 1239-1789, 1245-1489, 1290-1789, 1307-1798, 1308-1528, 1313-1798, 1321-1798, 1343-1758, 1350-1756, 1353-1799, 1374-1720, 1384-1789, 1389-1798, 1392-1666, 1392-1750, 1394-1637, 1395-1789, 1396-1798, 1404-1676, 1406-1798, 1414-1789, 1414-1799, 1420-1620, 1424-1703, 1444-1789, 1444-1799, 1448-1780, 1450-1750, 1453-1798, 1457-1732, 1457-1752, 1462-1676, 1476-1708, 1478-1780, 1493-1798, 1494-1771, 1494-1789, 1500-1798, 1524-1797,
	1527-1898, 1594-1867, 1607-1796, 1610-1798, 1637-1950, 1698-1798, 1717-1770, 1766-2184, 1791-2068, 1791-2187, 1791-2191, 1791-2192, 1791-2193, 1791-2197, 1791-2207, 1791-2218, 1792-2096, 1800-2134, 1800-2137, 1800-2153, 1800-2245, 1800-2256, 1801-2253, 1809-2188, 1815-2191, 1815-2277, 1818-2318, 1818-2320, 1824-2209, 1835-2103, 1836-2099, 1838-2336, 1846-2340, 1857-2106, 1857-2328, 1857-2361, 1942-2340, 1958-2338, 1968-2347, 2003-2239, 2020-2344, 2093-2340, 2106-2338, 2149-2350, 2154-2351, 2158-2398, 2187-2354
46/140052CB1/2127	1-614, 3-596, 5-642, 23-576, 43-639, 184-640, 220-640, 245-640, 373-628, 376-992, 377-1066, 378-562, 392-640, 421-689, 505-796, 505-1052, 524-1069, 532-619, 538-779, 595-1241, 600-1087, 606-814, 668-915, 679-1239, 685-991, 692-1145, 755-1352, 756-1035, 768-1052, 851-1113, 851-1354, 898-1231, 909-1412, 911-1126, 994-1280, 1012-1223, 1067-1338, 1074-1633, 1089-1523, 1125-1414, 1183-1698, 1192-1453, 1233-1518, 1234-1709, 1257-1856, 1258-1703, 1273-1836, 1321-1555, 1321-1841, 1342-1617, 1345-1486, 1357-1798, 1394-1575, 1454-1853, 1470-1843, 1535-1815, 1557-2119, 1580-1822, 1580-2056, 1583-1844, 1595-1870, 1619-1853, 1622-2074, 1623-2092, 1633-1921, 1640-2100, 1642-2101, 1642-2127, 1650-1873, 1650-1982, 1650-2067, 1651-1819, 1659-1910, 1668-2119, 1669-2112, 1671-1865, 1673-2074, 1686-2118, 1704-1988, 1723-2063, 1815-2100, 1867-2087, 1867-2114, 1897-2119, 1904-2048

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
47/5158048CB1/2407	1-311, 1-489, 8-257, 10-290, 13-239, 17-295, 22-333, 28-289, 29-308, 33-282, 51-294, 58-306, 58-311, 59-283, 92-336, 92-542, 216-414, 264-430, 328-548, 356-646, 489-711, 489-736, 505-613, 505-1105, 551-639, 577-872, 638-677, 638-701, 638-721, 638-736, 638-744, 638-754, 638-756, 638-765, 638-777, 638-786, 638-789, 638-809, 638-832, 638-833, 638-836, 638-879, 638-1040, 648-904, 652-897, 676-1049, 689-1049, 699-908, 699-1193, 708-914, 800-1069, 803-1072, 808-1074, 841-1049, 841-1214, 861-1151, 868-1032, 900-1100, 900-1448, 921-1193, 952-1199, 968-1186, 989-1213, 1014-1214, 1071-1214, 1075-1214, 1086-1296, 1086-1622, 1095-1214, 1153-1412, 1175-1469, 1212-1707, 1251-1486, 1266-1507, 1297-1352, 1297-1423, 1297-1426, 1297-1486, 1306-1564, 1308-1577, 1311-1553, 1318-1536, 1318-1540, 1343-1594, 1344-1601, 1351-1617, 1373-1664, 1374-1626, 1384-1642, 1391-1486, 1400-1629, 1429-1693, 1444-1717, 1464-1711, 1475-1774, 1477-1713, 1496-1784, 1522-1976, 1531-2018, 1535-1825, 1538-1729, 1565-1845, 1565-2006, 1567-1803, 1567-1867, 1575-2066, 1576-1872, 1576-1873, 1615-1901, 1622-1835, 1638-1684, 1638-1791, 1638-1804, 1645-1898, 1668-1926, 1671-1940, 1687-2233, 1715-
	2013, 1726-2387, 1727-1837, 1727-1859, 1727-1912, 1727-1945, 1727-1955, 1729-1970, 1729-2273, 1734-2389, 1739-1970, 1752-2377, 1752-2387, 1755-2016, 1755-2022, 1755-2024, 1755-2045, 1755-2363, 1757-2391, 1760-2379, 1763-2026, 1765-2381, 1765-2393, 1766-2390, 1775-2051, 1788-1934, 1788-2381, 1791-2037, 1795-2093, 1799-2377, 1808-2025, 1811-2078, 1811-2293, 1816-2084, 1820-2379, 1823-2316, 1827-2027, 1827-2290, 1828-2377, 1829-2301, 1834-2099, 1835-2096, 1836-2391, 1846-2377, 1857-2105, 1858-2122, 1858-2201, 1864-2068, 1864-2128, 1866-2406, 1867-1988, 1867-2092, 1867-2118, 1867-2387, 1877-2140, 1881-2123, 1893-2369, 1893-2382, 1895-2165, 1897-2112, 1897-2373, 1897-2407, 1906-2407, 1907-2173, 1911-2377, 1931-2216, 1932-2160, 1937-2172, 1944-2396, 1947-2172, 1949-2293, 1950-2147, 1950-2162, 1951-2215, 1951-2387, 1952-2396, 1953-2405, 1955-2387, 1964-2394, 1968-2407, 1972-2407, 1979-2407, 1982-2253, 1983-2397, 2004-2393, 2010-2189, 2010-2387, 2010-2407, 2020-2384, 2028-2392, 2029-2399, 2032-2391, 2037-2396, 2039-2401, 2048-2290, 2058-2240, 2065-2392, 2076-2345, 2081-2389, 2093-2157, 2093-2368, 2093-2384, 2099-2347, 2099-2366, 2099-2398,
	2106-2361, 2110-2319, 2115-2358, 2117-2390, 2124-2389, 2124-2404, 2126-2363, 2126-2398, 2134-2387, 2138-2373, 2140-2387, 2151-2407, 2167-2398, 2186-2398, 2189-2407, 2202-2407, 2213-2407, 2214-2407, 2216-2398, 2233-2407, 2235-2397, 2280-2398, 2315-2407, 2328-2390

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
48/3127541CB1/4549	1-487, 1-499, 1-549, 1-812, 3-367, 28-474, 29-482, 49-337, 88-437, 217-732, 289-525, 289-776, 329-619, 406-823, 419-711, 550-1119, 753-1048, 851-1530, 948-1187, 948-3967, 1028-1187, 1028-1309, 1075-1679, 1125-1726, 1188-1375, 1310-1519, 1552-1994, 1631-1850, 1730-1850, 1730-1993, 1810-2114, 1851-1993, 1851-2114, 1952-2220, 2117-2284, 2165-2655, 2186-2655, 2286-3967, 3016-3531, 3016-3557, 3016-3565, 3016-3729, 3297-4008, 3309-4008, 3381-4019, 3381-4032, 3407-3977, 3534-3801, 3590-3921, 3601-3921, 3642-3921, 3647-3921, 3656-3921, 3662-3929, 3788-4549
49/8224777CB1/2598	1-558, 347-600, 451-558, 493-852, 559-588, 559-852, 772-1122, 773-1121, 775-820, 922-1290, 922-1338, 926-1338, 1170-1297, 1239-1769, 1417-1600, 1495-1603, 1536-1570, 1536-1600, 1536-1626, 1536-1740, 1536-1761, 1536-1781, 1536-1818, 1678-2016, 1817-2334, 1817-2410, 1823-2464, 1860-2581, 1878-2506, 1888-2598
50/587394CB1/1353	1-655, 191-655, 197-655, 199-655, 214-655, 218-701, 219-701, 221-653, 228-655, 230-682, 237-701, 239-682, 253-701, 264-655, 265-655, 266-654, 275-698, 277-678, 281-682, 284-655, 301-655, 304-669, 322-655, 353-653, 362-654, 378-650, 395-648, 421-682, 421-867, 431-655, 436-701, 465-740, 478-726, 509-787, 514-1353, 622-890, 690-969, 766-1220
51/1402405CB1/2161	1-615, 1-735, 385-628, 464-1088, 493-2003, 501-881, 503-931, 515-772, 945-1199, 975-1235, 979-1258, 979-1387, 1073-1317, 1080-1328, 1100-1329, 1148-1387, 1157-1387, 1183-1500, 1365-2036, 1378-1657, 1438-1985, 1461-1705, 1466-2062, 1484-1722, 1484-2055, 1486-1684, 1486-1703, 1526-2121, 1527-2039, 1545-1744, 1555-2111, 1595-1820, 1604-1892, 1642-2111, 1664-2159, 1669-2111, 1722-1994, 1730-1980, 1783-2039, 1796-2139, 1811-2161, 1864-2073, 1930-2062
52/1798468CB1/1487	1-278, 1-472, 17-293, 17-573, 19-295, 20-152, 20-268, 27-265, 27-418, 31-293, 35-426, 50-450, 53-224, 55-326, 61-634, 62-589, 63-284, 63-302, 63-304, 63-420, 64-298, 65-404, 68-312, 69-341, 96-796, 112-261, 301-721, 306-800, 312-800, 338-727, 338-755, 342-790, 343-807, 346-720, 353-742, 385-805, 386-805, 396-805, 398-788, 402-717, 403-744, 408-747, 421-717, 429-805, 437-810, 454-801, 460-812, 478-718, 552-800, 592-743, 630-1057, 806-1290, 816-1343, 816-1487

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
53/3189084CB1/2495	1-331, 32-146, 35-604, 115-183, 218-767, 231-836, 304-739, 334-846, 395-628, 454-960, 463-951, 495-774, 507-1090, 508-875, 511-852, 512-798, 514-1176, 565-657, 586-864, 609-1274, 615-940, 615-1222, 618-1355, 621-949, 648-1038, 649-1023, 686-996, 703-1267, 761-1373, 776-1147, 810-1382, 845-874, 913-1469, 920-1238, 920-1384, 944-1555, 983-1356, 985-1586, 990-1465, 994-1088, 1000-1305, 1015-1249, 1069-1591, 1116-1405, 1117-1699, 1121-1707, 1162-1416, 1170-1798, 1337-1656, 1343-1919, 1363-1621, 1378-1624, 1415-1598, 1415-1945, 1472-1733, 1486-2127, 1499-1755, 1519-2092, 1542-1814, 1555-1828, 1572-2175, 1622-2118, 1636-1883, 1703-2014, 1716-2278, 1750-2010, 1862-2277, 1914-2141, 1914-2446, 1914-2495, 1924-2374, 1934-2204, 2001-2258, 2020-2294, 2021-2285, 2098-2411, 2377-2494
54/5580384CB1/2227	1-65, 1-257, 1-595, 1-611, 5-159, 6-273, 7-244, 7-263, 8-244, 8-483, 9-330, 9-481, 10-316, 10-322, 11-335, 13-251, 14-268, 15-159, 15-192, 15-426, 20-271, 20-319, 20-518, 23-317, 30-253, 30-600, 30-624, 33-273, 33-325, 36-265, 44-523, 46-343, 54-229, 71-435, 120-451, 128-426, 136-687, 257-687, 258-687, 270-728, 275-1237, 276-728, 290-733, 297-728, 341-711, 362-730, 368-730, 413-705, 435-732, 442-961, 446-693, 446-743, 495-694, 495-711, 521-834, 522-705, 526-728, 526-823, 557-816, 558-728, 588-728, 599-994, 626-835, 762-1051, 775-1023, 776-1042, 776-1289, 811-1087, 840-1056, 850-1079, 921-1193, 970-1233, 970-1246, 982-1240, 985-1152, 1024-1247, 1044-1483, 1098-1314, 1104-1354, 1129-1248, 1143-1609, 1171-1420, 1171-1468, 1180-1392, 1181-1469, 1181-1675, 1202-1512, 1210-1442, 1224-1457, 1224-1466, 1271-1432, 1272-1548, 1272-1799, 1278-1512, 1288-1434, 1288-1496, 1298-1581, 1303-1916, 1305-1914, 1316-1532, 1316-1555, 1316-1915, 1330-1565, 1336-1922, 1378-1641, 1384-1684, 1417-1853, 1418-1613, 1438-1934, 1442-1691, 1471-1655, 1480-1743, 1480-1992, 1482-1681, 1518-1796, 1519-1963, 1526-1766, 1528-1953, 1530-1794, 1533-1811, 1549-1960, 1559-1775, 1572-1960, 1575-1956,
	1592-2104, 1593-1955, 1598-1956, 1620-1800, 1633-2077, 1644-1946, 1674-1923, 1696-1960, 1707-1962, 1723-1963, 1972-2227
55/5158619CB1/1652	1-277, 1-526, 5-485, 10-294, 12-258, 15-522, 17-273, 17-595, 376-410, 479-758, 511-775, 512-767, 512-1023, 607-903, 607-1077, 739-1445, 804-1009, 804-1331, 810-1306, 837-1135, 837-1362, 951-1556, 992-1613, 1022-1619, 1033-1618, 1049-1492, 1064-1233, 1065-1269, 1065-1301, 1065-1603, 1065-1627, 1165-1606, 1165-1644, 1185-1437, 1185-1564, 1188-1449, 1204-1638, 1208-1483, 1228-1627, 1286-1417, 1286-1516, 1286-1549, 1286-1598, 1385-1648, 1398-1646, 1456-1642, 1457-1543, 1557-1624, 1577-1652
56/2792745CB1/696	1-696

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
57/2827678CB1/600	1-600, 35-93, 205-263, 310-433, 400-446, 438-562
58/790257CB1/3005	1-1403, 426-1700, 433-712, 451-1243, 456-999, 461-1054, 472-1030, 473-1048, 474-679, 474-928, 482-1151, 494-730, 495-761, 500-756, 507-1183, 514-963, 514-1174, 517-747, 524-1265, 542-912, 550-917, 567-761, 609-1247, 686-1313, 709-986, 709-1151, 822-1254, 889-1185, 1050-1677, 1107-1379, 1111-1780, 1271-1889, 1358-1559, 1381-2143, 1458-1686, 1488-1741, 1496-2043, 1499-2118, 1505-1941, 1520-1800, 1550-1774, 1588-1875, 1632-2003, 1635-2105, 1651-1950, 1662-1973, 1663-1963, 1681-2249, 1722-1875, 1746-2008, 1752-2464, 1760-1898, 1812-2118, 1819-2151, 1836-2118, 1856-2212, 1879-2265, 1884-2554, 1892-2534, 1963-2404, 1970-2209, 1970-2492, 2010-2291, 2020-2267, 2036-2345, 2042-2317, 2077-2327, 2090-2753, 2105-2754, 2128-2420, 2128-2454, 2129-2257, 2145-2724, 2156-2693, 2158-2411, 2176-2336, 2192-2788, 2193-2720, 2209-2780, 2258-2703, 2314-2777, 2314-2817, 2319-2800, 2343-2800, 2346-2800, 2356-2800, 2360-2809, 2361-2817, 2364-2817, 2371-2800, 2374-2868, 2381-2800, 2385-2802, 2386-2804, 2389-2800, 2390-2802, 2402-2662, 2402-2664, 2405-2628, 2405-
	2903, 2405-2907, 2415-2676, 2415-2919, 2416-2959, 2421-2945, 2422-2803, 2425-2802, 2460-2800, 2477-2802, 2483-2802, 2489-2806, 2490-2639, 2490-2967, 2509-2959, 2514-2968, 2516-2800, 2520-2963, 2526-2971, 2527-2800, 2546-2968, 2588-2815, 2595-2957, 2624-2963, 2625-2962, 2638-2963, 2684-2817, 2687-2923, 2731-2968, 2822-2971, 2831-2970, 2849-2968, 2850-3005
59/2617345CB1/2530	1-250, 1-558, 28-260, 38-334, 39-266, 40-261, 40-288, 45-560, 45-566, 51-271, 52-344, 88-395, 262-549, 271-451, 285-577, 293-476, 300-899, 321-564, 383-638, 391-962, 398-637, 415-715, 420-698, 420-947, 422-669, 461-935, 483-748, 541-578, 554-581, 568-833, 578-868, 625-761, 657-897, 685-901, 722-989, 750-993, 752-999, 752-1008, 786-1054, 786-1206, 794-1085, 808-1058, 813-1276, 835-1123, 844-1291, 857-1101, 858-1104, 882-1212, 897-1199, 901-1255, 914-1129, 946-1157, 946-1202, 947-1345, 948-1345, 957-1063, 987-1247, 1005-1257, 1017-1296, 1037-1345, 1041-1268, 1041-1277, 1041-1296, 1079-1331, 1085-1336, 1085-1340, 1100-1360, 1113-1310, 1173-1418, 1184-1441, 1205-1461, 1274-1456, 1290-1537, 1350-1596, 1350-1602, 1370-1618, 1397-1634, 1425-1707, 1432-1686, 1432-1739, 1436-1701, 1476-1732, 1512-1693, 1523-1753, 1542-1786, 1571-1838, 1612-1847, 1636-1880, 1636-1905, 1640-1858, 1653-1967, 1675-1928, 1706-1927, 1762-2019, 1762-2025, 1777-2035, 1780-2351, 1805-2491, 1809-2013, 1826-2494, 1847-2042, 1847-2049, 1916-2137, 1940-2128, 1958-2208, 1959-2159, 1959-2202, 1959-2516, 1961-2492, 1961-2530, 1983-2261, 2073-2310, 2081-2347, 2098-2317, 2119-2369, 2165-2499,
	2175-2440, 2235-2507, 2298-2507, 2300-2467, 2324-2507, 2357-2530

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
60/3254666CB1/1625	1-248, 90-310, 90-322, 101-395, 170-436, 170-736, 243-685, 250-542, 253-733, 275-732, 278-729, 279-566, 279-732, 311-731, 318-736, 335-575, 335-718, 335-729, 335-731, 378-731, 435-642, 509-731, 551-731, 589-832, 642-1001, 743-1015, 745-1260, 886-1532, 933-1233, 1064-1525, 1122-1435, 1149-1438, 1149-1560, 1149-1625
61/4159378CB1/1795	1-74, 1-711, 72-243, 80-639, 116-879, 118-377, 129-651, 133-375, 148-449, 242-752, 271-519, 305-935, 421-994, 423-596, 561-856, 561-1205, 565-866, 817-1320, 900-1504, 925-1209, 949-1219, 949-1370, 1015-1286, 1015-1289, 1016-1253, 1027-1263, 1027-1522, 1113-1722, 1124-1720, 1136-1440, 1237-1777, 1244-1492, 1245-1674, 1262-1784, 1263-1719, 1292-1780, 1312-1762, 1327-1577, 1331-1788, 1334-1771, 1347-1787, 1357-1795, 1385-1762, 1421-1689, 1437-1795, 1449-1795, 1461-1774, 1461-1780, 1461-1790, 1522-1788, 1537-1788
62/4317538CB1/2080	1-341, 1-620, 1-621, 1-647, 1-701, 2-658, 208-755, 217-749, 221-755, 346-581, 453-1146, 457-1146, 471-1146, 502-1146, 581-1355, 613-1146, 617-1328, 662-1345, 664-1452, 670-1365, 950-1577, 972-1648, 1027-1723, 1040-1510, 1043-1698, 1046-1842, 1053-1752, 1091-1729, 1177-1897, 1197-2080, 1204-1789, 1227-1877, 1243-1937, 1245-1936, 1267-1849
63/1881010CB1/1599	1-561, 2-155, 2-610, 6-569, 6-786, 8-271, 36-642, 38-317, 53-571, 58-157, 59-758, 60-347, 66-766, 69-488, 97-598, 168-446, 252-1139, 377-962, 392-664, 398-924, 427-1019, 467-1236, 485-1159, 496-860, 497-770, 532-821, 548-1178, 569-824, 603-1177, 604-691, 630-1322, 634-851, 642-1200, 652-1302, 670-919, 677-1293, 708-966, 711-1261, 711-1271, 714-1203, 722-1307, 750-1444, 765-1455, 815-1297, 815-1443, 834-1298, 856-1493, 857-1369, 861-1096, 877-1118, 884-1558, 887-1115, 887-1437, 903-1439, 905-1575, 907-1305, 920-1115, 921-1476, 929-1154, 954-1238, 958-1566, 963-1526, 966-1411, 967-1175, 977-1230, 977-1565, 982-1250, 987-1224, 987-1599, 989-1381, 991-1256, 1018-1266, 1019-1528, 1019-1573, 1023-1379, 1044-1316, 1045-1377, 1050-1294, 1059-1333, 1061-1325, 1070-1330, 1088-1329, 1269-1550

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
64/1593038CB1/4137	1-465, 1-587, 26-355, 361-1067, 387-897, 427-563, 499-921, 632-1283, 642-921, 652-1115, 690-942, 778-920, 781-1127, 793-1362, 798-1414, 881-1031, 922-1201, 928-1356, 934-1436, 983-1265, 1007-1202, 1007-1225, 1069-1799, 1109-1782, 1124-1334, 1125-1783, 1125-1921, 1125-1932, 1141-1781, 1145-1748, 1146-1711, 1146-1744, 1173-1448, 1179-1412, 1185-1870, 1190-1404, 1225-1471, 1271-4022, 1274-1893, 1285-2076, 1291-1932, 1317-2024, 1354-2033, 1385-2205, 1499-1812, 1504-1872, 1511-1808, 1511-1811, 1511-1826, 1541-1861, 1541-2285, 1627-1811, 1711-2152, 1714-2492, 1792-1980, 1812-2496, 1822-2069, 1822-2436, 1869-2131, 1871-2154, 1877-2172, 1910-2569, 1924-2290, 1936-2623, 1941-2162, 1948-2576, 1958-2193, 1962-2474, 1963-2132, 1963-2380, 1981-2605, 2005-2196, 2051-2262, 2054-2715, 2058-2301, 2058-2556, 2067-2652, 2079-2352, 2098-2547, 2100-2236, 2100-2702, 2112-2399, 2121-2766, 2141-2662, 2161-2425, 2162-2405, 2164-2363, 2183-2456, 2188-2424, 2195-2431, 2195-2463, 2198-2810, 2203-2796, 2208-2888, 2213-2554, 2219-2873, 2228-2827, 2262-2925, 2368-
	2929, 2385-2844, 2476-3069, 2480-2928, 2484-3058, 2488-3155, 2517-2800, 2531-3103, 2531-3108, 2537-2844, 2558-2772, 2558-3159, 2560-2804, 2560-2843, 2560-2878, 2560-2887, 2560-3157, 2560-3248, 2562-3065, 2563-2945, 2563-3070, 2563-3112, 2563-3206, 2563-3222, 2563-3243, 2563-3272, 2565-2803, 2569-3127, 2573-3121, 2608-2711, 2619-2888, 2634-3166, 2643-3262, 2644-3122, 2644-3138, 2644-3236, 2651-3340, 2653-3151, 2661-3332, 2663-3284, 2664-3206, 2664-3327, 2666-3288, 2670-2971, 2672-2865, 2681-2904, 2693-3285, 2694-2978, 2699-3366, 2705-3121, 2705-3177, 2710-2899, 2724-3363, 2724-3376, 2725-3387, 2730-2936, 2731-3307, 2735-3418, 2746-3310, 2747-3023, 2756-3453, 2763-3019, 2774-2940, 2774-3127, 2780-3217, 2786-3043, 2802-3353, 2813-3402, 2814-3341, 2816-3340, 2817-3336, 2818-3360, 2825-3326, 2833-3580, 2834-3405, 2834-3547, 2848-3467, 2857-3514, 2858-3345, 2866-3385, 2867-3536, 2880-3554, 2888-3369, 2890-3323, 2894-3460, 2918-3531, 2929-3393, 2932-3471, 2936-3610, 2947-3474, 2954-3101, 2955-3630, 2975-3519, 2976-3479, 2991-3560, 3014-3689, 3016-3555, 3018-3585, 3034-3605, 3035-3573, 3036-3326, 3041-3525, 3042-3725, 3046-3362, 3054-3333,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
64 (cont.)	3058-3725, 3065-3585, 3068-3725, 3069-3337, 3069-3432, 3073-3561, 3073-3725, 3074-3344, 3075-3725, 3076-3725, 3077-3725, 3078-3725, 3079-3514, 3080-3725, 3084-3563, 3084-3725, 3086-3725, 3087-3725, 3089-3725, 3095-3539, 3098-3725, 3099-3293, 3104-3647, 3104-3656, 3104-3696, 3113-3725, 3114-3725, 3119-3688, 3120-3725, 3125-3405, 3125-3683, 3127-3836, 3133-3725, 3134-3725, 3137-3347, 3139-3835, 3140-3467, 3140-3501, 3140-3725, 3140-3738, 3141-3725, 3142-3725, 3145-3725, 3146-3389, 3147-3406, 3148-3784, 3149-3725, 3151-3725, 3151-3755, 3152-3614, 3152-3718, 3152-3725, 3156-3725, 3159-3442, 3159-3725, 3159-3731, 3159-3799, 3163-3725, 3164-3725, 3164-3732, 3165-3725, 3165-3742, 3166-3725, 3167-3725, 3169-3725, 3172-3725, 3174-3440, 3174-3819, 3176-3596, 3176-3725, 3177-3470, 3177-3725, 3180-3725, 3181-3533, 3183-3427, 3183-3703, 3183-3725, 3183-3751, 3192-3580, 3192-3662, 3193-3825, 3195-3956, 3200-3725, 3202-3725, 3208-3725, 3208-3918, 3213-3725, 3214-3725, 3217-3356, 3218-3725, 3219-3725, 3220-3565, 3226-3725, 3230-3864, 3231-3725, 3235-3725, 3236-3725, 3237-3537, 3249-3774, 3251-3554, 3253-3653, 3253-3673, 3257-3500, 3260-3400, 3264-
	3510, 3266-3731, 3267-3725, 3269-3725, 3273-3531, 3275-3725, 3284-3389, 3286-3978, 3288-3725, 3291-3556, 3293-3552, 3293-3757, 3293-3879, 3300-3881, 3300-3927, 3313-3515, 3316-3725, 3318-3587, 3318-3725, 3319-3618, 3321-3716, 3323-3721, 3323-3901, 3328-3609, 3328-3669, 3332-3620, 3334-3725, 3341-3626, 3341-3652, 3343-3595, 3343-3837, 3348-3873, 3350-3947, 3355-3605, 3356-3725, 3360-3570, 3367-3733, 3368-3624, 3369-4083, 3370-3629, 3371-3624, 3371-3725, 3372-3549, 3373-3791, 3373-4137, 3374-3617, 3374-3659, 3376-4026, 3380-3672, 3384-3958, 3386-3605, 3387-3930, 3388-4013, 3389-3891, 3391-3643, 3391-3848, 3397-3605, 3401-3887, 3407-3725, 3409-3955, 3604-3731, 3688-3725, 4040-4109
65/7494930CB1/4019	1-1240, 1-4019
66/7497349CB1/1965	1-548, 1-1965, 86-775, 89-775, 90-775, 91-774, 95-775, 106-775, 179-775, 184-775, 326-775, 462-750, 490-809, 708-807, 765-1262, 785-1410, 836-1433, 1276-1872, 1282-1965
67/5510805CB1/499	1-496, 1-499, 220-499

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
68/1577482CB1/3760	1-831, 130-399, 409-859, 410-813, 410-874, 410-914, 413-726, 413-810, 414-791, 414-847, 414-864, 415-449, 417-842, 417-865, 418-549, 418-685, 418-890, 418-899, 418-900, 429-946, 433-785, 433-912, 445-857, 448-773, 456-795, 460-918, 462-675, 462-749, 462-849, 462-906, 462-1083, 464-752, 464-868, 469-562, 475-847, 478-848, 483-923, 484-1160, 540-1011, 596-1008, 596-1212, 678-1305, 822-1024, 852-1019, 904-1497, 922-1538, 959-1619, 970-1490, 1053-1327, 1053-1661, 1057-1496, 1061-1496, 1063-1509, 1089-1261, 1089-1493, 1108-1523, 1111-1438, 1119-1743, 1158-1501, 1166-1877, 1183-1528, 1238-1929, 1250-1682, 1342-1443, 1342-1899, 1343-1447, 1343-1531, 1348-1956, 1349-1579, 1352-1837, 1352-1875, 1364-1930, 1368-1717, 1379-1681, 1379-2078, 1384-1736, 1386-1962, 1406-1946, 1413-1934, 1414-1743, 1458-1603, 1472-2108, 1488-1591, 1488-1730, 1488-2002, 1488-2055, 1556-2187, 1575-2007, 1593-1851, 1598-1777, 1623-1862, 1644-1850, 1650-1889, 1702-1941, 1719-1936, 1719-2287, 1728-1956, 1733-2217, 1751-2014, 1758-2147, 1759-2247, 1820-2014, 1831-2063, 1838-2114, 1854-2122, 1899-2542, 1924-2559, 1942-2175, 1950-2500, 1968-2593, 1972-2241, 1997-2271, 1998-2259, 2003-
	2430, 2019-2514, 2029-2556, 2034-2385, 2048-2624, 2049-2856, 2075-2305, 2078-2576, 2080-2362, 2105-2609, 2114-2461, 2119-2762, 2121-2608, 2126-2607, 2132-2277, 2132-2411, 2135-2608, 2137-2373, 2142-2609, 2153-2612, 2154-2608, 2159-2596, 2167-2371, 2167-2608, 2172-2609, 2172-2612, 2173-2608, 2174-2608, 2178-2290, 2183-2613, 2185-2608, 2188-2779, 2189-2609, 2189-2764, 2202-2609, 2210-2608, 2231-2609, 2231-2610, 2231-2613, 2234-2880, 2236-2504, 2259-2438, 2264-2609, 2269-2779, 2288-2601, 2290-2608, 2291-2779, 2306-3075, 2317-2593, 2320-2713, 2320-2715, 2338-2695, 2339-2609, 2350-2586, 2358-2610, 2361-2598, 2361-2601, 2373-2401, 2393-2666, 2457-2716, 2473-2768, 2474-2608, 2480-2812, 2488-2796, 2568-3287, 2588-3002, 2599-3209, 2656-3074, 2694-3048, 2694-3124, 2701-3128, 2709-3058, 2721-2853, 2729-2996, 2733-3031, 2774-3086, 2775-3381, 2783-3031, 2783-3243, 2783-3362, 2783-3451, 2786-3072, 2799-3447, 2805-3074, 2819-3146, 2824-3019, 2827-3148, 2827-3370, 2859-3052, 2874-3251, 2881-3381, 2904-3144, 2907-3583, 2919-3469, 2923-3390, 2974-3067, 3001-3329, 3051-3388, 3053-3294, 3053-3554, 3071-3597, 3074-3751, 3093-3756, 3096-3465, 3125-3744,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
68 (cont.)	3126-3380, 3126-3461, 3130-3733, 3136-3365, 3165-3456, 3180-3476, 3180-3483, 3211-3745, 3213-3745, 3221-3760, 3260-3742, 3263-3733, 3266-3750, 3269-3742, 3281-3738, 3282-3760, 3284-3650, 3285-3742, 3293-3744, 3294-3751, 3298-3750, 3300-3539, 3300-3753, 3302-3742, 3307-3750, 3317-3751, 3321-3604, 3342-3750, 3345-3745, 3347-3752, 3353-3650, 3353-3752, 3362-3753, 3365-3756, 3372-3750, 3374-3759, 3398-3706, 3399-3754, 3400-3626, 3400-3760, 3402-3745, 3407-3751, 3418-3632, 3430-3745, 3431-3697, 3432-3745, 3439-3752, 3451-3744, 3463-3742, 3466-3730, 3466-3751, 3466-3756, 3473-3708, 3473-3713, 3473-3751, 3504-3639, 3504-3758, 3532-3744, 3544-3752, 3545-3752, 3570-3744, 3587-3741, 3611-3745, 3662-3747
69/1805054CB1/2800	1-316, 221-724, 221-741, 221-846, 221-862, 224-814, 225-863, 227-405, 230-507, 230-543, 231-824, 240-504, 240-824, 246-552, 250-482, 253-1037, 261-509, 266-533, 266-694, 283-516, 283-557, 299-574, 604-828, 617-1219, 618-1134, 618-1219, 618-1236, 618-1253, 619-879, 686-1299, 721-1245, 794-873, 843-1105, 867-1480, 894-1145, 952-1472, 961-1084, 1001-1671, 1038-1333, 1112-1344, 1112-1580, 1290-1556, 1463-1657, 1513-1773, 1525-1802, 1557-2126, 1568-2168, 1573-1838, 1573-2273, 1583-2031, 1603-2292, 1622-2220, 1630-2140, 1661-2220, 1680-1907, 1696-2339, 1717-1976, 1743-2039, 1753-2023, 1826-2351, 1830-2099, 1910-2093, 1927-2153, 1986-2757, 1998-2264, 1998-2586, 2000-2345, 2000-2414, 2000-2547, 2069-2303, 2076-2242, 2076-2260, 2076-2283, 2076-2284, 2076-2295, 2093-2340, 2096-2291, 2106-2165, 2107-2383, 2128-2766, 2129-2760, 2177-2769, 2185-2720, 2195-2438, 2195-2440, 2197-2763, 2204-2412, 2214-2442, 2224-2680, 2244-2541, 2248-2720, 2248-2772, 2250-2759, 2255-2437, 2279-2550, 2281-2783, 2388-2800, 2472-2754, 2501-2775, 2535-2778, 2551-2800
70/7492708CB1/1314	1-1314, 101-1314, 351-1314
71/7490847CB1/3594	1-278, 1-296, 204-873, 223-875, 237-296, 318-572, 318-589, 383-646, 600-2788, 766-1419, 768-1316, 956-1356, 977-1220, 977-1436, 977-1524, 980-1436, 1093-1561, 1152-1524, 1292-1436, 1589-1864, 1589-1949, 1589-1963, 1723-1944, 2271-2728, 2285-3005, 2327-2763, 2347-2770, 2411-2775, 2411-2776, 2413-2770, 2439-3103, 2442-2670, 2545-3102, 2699-2759, 2749-3189, 2786-3381, 2788-3381, 2828-3350, 2828-3424, 2893-3490, 2894-3051, 3117-3246, 3214-3487, 3232-3483, 3232-3583, 3232-3594
72/7493059CB1/4123	1-409, 1-1272, 1-1917, 1-2436, 1-3909, 1-4123

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
73/2321130CB1/2569	1-896, 54-957, 59-2064, 148-854, 180-837, 405-639, 406-662, 406-860, 406-882, 406-891, 410-706, 410-940, 414-1007, 415-707, 415-898, 415-900, 415-919, 415-932, 415-955, 415-961, 415-1043, 415-1058, 416-903, 416-1019, 416-1023, 416-1043, 435-521, 443-894, 448-691, 448-801, 448-908, 451-730, 456-903, 457-907, 470-975, 482-787, 491-852, 495-830, 617-975, 619-975, 632-1144, 657-975, 660-846, 660-867, 708-1412, 739-922, 747-1102, 749-1427, 758-1154, 925-1483, 930-1410, 943-1362, 1045-1118, 1062-1720, 1064-1388, 1067-1649, 1083-1317, 1104-1687, 1188-1752, 1195-1739, 1285-1853, 1300-1546, 1309-1843, 1334-1410, 1348-2028, 1390-1821, 1392-2045, 1415-1441, 1425-1664, 1425-1856, 1479-2296, 1586-2222, 1605-2259, 1676-2185, 1698-2064, 1709-2013, 1709-2569, 1710-2255, 1712-2204, 1714-2463, 1717-2096, 1724-2039, 1738-1980, 1739-2283, 1770-2247, 1774-2246, 1776-2360, 1794-1959, 1794-2142, 1803-2234, 1892-2207, 2460-2488, 2460-2516, 2460-2537, 2463-2516
74/2008365CB1/1066	1-259, 1-636, 120-415, 128-464, 303-807, 303-835, 305-737, 360-835, 363-830, 365-837, 367-837, 369-827, 369-835, 374-827, 375-827, 377-837, 384-837, 393-831, 426-836, 428-837, 431-831, 431-837, 436-827, 446-827, 450-720, 451-835, 477-835, 480-835, 486-598, 486-824, 490-836, 493-835, 533-830, 537-827, 581-833, 586-813, 586-828, 586-831, 632-1066, 730-837
75/3580778CB1/1817	1-337, 1-1817, 219-838, 219-1013, 240-864, 240-1044, 254-346, 255-1812, 296-444, 298-579, 337-1532, 369-1316, 369-1351, 529-1279, 530-1135, 530-1407, 599-1405, 602-1405, 643-1405, 755-1405, 853-1467, 870-1463, 881-1310, 896-1304, 898-1310, 1050-1513, 1065-1535, 1141-1541
76/7948785CB1/2391	1-706, 565-1646, 816-1284, 816-1293, 816-1443, 816-1474, 816-1482, 816-1494, 818-1064, 820-1003, 820-1175, 820-1221, 820-1238, 820-1272, 820-1355, 820-1362, 820-1370, 820-1388, 820-1408, 820-1428, 820-1438, 820-1446, 820-1518, 820-1520, 820-1556, 820-1566, 820-1624, 820-1628, 820-1638, 820-1645, 821-1528, 821-1567, 822-1522, 822-1612, 823-1405, 823-1462, 823-1480, 823-1615, 828-1623, 837-1634, 851-1617, 861-1533, 877-1698, 1005-1385, 1010-1431, 1010-1528, 1106-1876, 1123-1463, 1142-1474, 1391-2000, 1416-1663, 1417-2080, 1475-2391, 1502-1879, 1510-2309
77/7494415CB1/1314	1-1314, 101-1314
78/2234223CB1/2076	1-49, 1-244, 1-450, 1-555, 1-567, 1-576, 1-587, 1-615, 1-631, 1-643, 1-659, 1-2066, 217-869, 557-1441, 563-1296, 937-1533, 1024-1713, 1034-1519, 1128-1661, 1146-1773, 1147-1905, 1163-1870, 1197-1876, 1386-2076, 1427-2076, 1434-2058, 1765-2076

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
40	4973222CB1	MPHGLPT02
41	55009060CB1	UCMCL5T01
42	1985092CB1	LUNGNOT12
43	1553593CB1	KIDNNOT05
44	1954122CB1	SPLNNOT04
45	3159276CB1	THP1AZS08
46	140052CB1	MYOMNOT01
47	5158048CB1	BRSTTUT03
48	3127541CB1	PLACNOR01
49	8224777CB1	MUSLTDR02
50	587394CB1	UTRSNOT01
51	1402405CB1	KIDNNOT05
52	1798468CB1	SINTFER02
53	3189084CB1	BONEUNR01
54	5580384CB1	ADRENOT07
55	5158619CB1	MUSCNOT07
58	790257CB1	PROSTUT03
59	2617345CB1	GBLANOT01
60	3254666CB1	UTRSNOR01
61	4159378CB1	ISLTNOT01
62	4317538CB1	BRAIFEE05
63	1881010CB1	STOMTDA01
64	1593038CB1	SKINDIA01
66	7497349CB1	BRAGNON02
67	5510805CB1	BRADDIR01
68	1577482CB1	BRAUNOR01
69	1805054CB1	PLACNOB01
71	7490847CB1	BRAIFEE05
73	2321130CB1	KIDEUNE02
74	2008365CB1	TESTNOT03
75	3580778CB1	THYMDIT01
76	7948785CB1	BRAWTDR02
78	2234223CB1	PANCTUT02

Table 6

Library	Vector	Library Description
ADRENOT07	pINCY	Library was constructed using RNA isolated from adrenal tissue removed from a 61-year-old female during a bilateral adrenalectomy. Patient history included an unspecified disorder of the adrenal glands.
BONEUNR01	PCDNA2.1	This random primed library was constructed using pooled cDNA from two different donors. cDNA was generated using mRNA isolated from an untreated MG-63 cell line derived from an osteosarcoma tumor removed from a 14-year-old Caucasian male (donor A) and using mRNA isolated from sacral bone tumor tissue removed from an 18-year-old Caucasian female (donor B) during an exploratory laparotomy and soft tissue excision. Pathology indicated giant cell tumor of the sacrum in donor B. Donor B's history included pelvic joint pain, constipation, urinary incontinence, unspecified abdominal/pelvic symptoms, and a pelvic soft tissue malignant neoplasm. Family history included prostate cancer in donor B.
BRADDIR01	pINCY	Library was constructed using RNA isolated from diseased choroid plexus tissue of the lateral ventricle, removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident.
BRAGNON02	pINCY	This normalized substantia nigra tissue library was constructed from 4.2 10e7 independent clones from a substantia nigra tissue library. Starting RNA was made from RNA isolated from substantia nigra tissue removed from an 81-year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids, bilaterally; microscopic infarcts of the frontal cortex and hippocampus; and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. Grossly, the leptomeninges showed only mild thickening and hyalinization along the superior sagittal sinus. The remainder of the leptomeninges was thin and contained some congested blood vessels. Mild atrophy was found mostly in the frontal poles and lobes, and temporal lobes, bilaterally. Microscopically, there were pairs of Alzheimer type II astrocytes within the deep layers of the neocortex. There was increased satellitosis around neurons in the deep gray matter in the middle frontal cortex. The amygdala contained rare diffuse plaques and neurofibrillary tangles. The posterior hippocampus
		contained a microscopic area of cystic cavitation with hemosiderin-laden macrophages surrounded by reactive gliosis. Patient history included sepsis, cholangitis, post-operative atelectasis, pneumonia CAD, cardiomegaly due to left ventricular hypertrophy, splenomegaly, arteriolonephrosclerosis, nodular colloidal goiter, emphysema, CHF, hypothyroidism, and peripheral vascular disease. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRAIFEE05	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation.

Table 6

Library	Vector	Library Description
BRAUNOR01	pINCY	This random primed library was constructed using RNA isolated from striatum, globus pallidus and posterior putamen tissue removed from an 81-year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids, bilaterally; microscopic infarcts of the frontal cortex and hippocampus; and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. Grossly, the leptomeninges showed only mild thickening and hyalinization along the superior sagittal sinus. The remainder of the leptomeninges was thin and contained some congested blood vessels. Mild atrophy was found mostly in the frontal poles and lobes, and temporal lobes, bilaterally. Microscopically, there were pairs of Alzheimer type II astrocytes within the deep layers of the neocortex. There was increased satellitosis around neurons in the deep gray matter in the middle frontal cortex. The amygdala contained rare diffuse plaques and neurofibrillary tangles. The posterior hippocampus contained a microscopic area of cystic cavitation with hemosiderin-laden macrophages
		surrounded by reactive gliosis. Patient history included sepsis, cholangitis, post-operative atelectasis, pneumonia CAD, cardiomegaly due to left ventricular hypertrophy, splenomegaly, arteriolonephrosclerosis, nodular colloidal goiter, emphysema, CHF, hypothyroidism, and peripheral vascular disease.
BRAWTDR02	PCDNA2.1	This random primed library was constructed using RNA isolated from dentate nucleus tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRSTTUT03	PSPORT1	Library was constructed using RNA isolated from breast tumor tissue removed from a 58-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated multicentric invasive grade 4 lobular carcinoma. The mass was identified in the upper outer quadrant, and three separate nodules were found in the lower outer quadrant of the left breast. Patient history included skin cancer, rheumatic heart disease, osteoarthritis, and tuberculosis. Family history included cerebrovascular disease, coronary artery aneurysm, breast cancer, prostate cancer, atherosclerotic coronary artery disease, and type I diabetes.
GBLANOT01	pINCY	Library was constructed using RNA isolated from diseased gallbladder tissue removed from a 53-year-old Caucasian female during a cholecystectomy. Pathology indicated mild chronic cholecystitis and cholelithiasis with approximately 150 mixed gallstones. Family history included benign hypertension.

Table 6

Library	Vector	Library Description
ISLTNOT01	pINCY	Library was constructed using RNA isolated from a pooled collection of pancreatic islet cells.
KIDEUNE02	pINCY	This 5' biased random primed library was constructed using RNA isolated from an untreated transformed embryonal cell line (293-EBNA) derived from kidney epithelial tissue (Invitrogen). The cells were transformed with adenovirus 5 DNA.
KIDNNOT05	PSPORT1	Library was constructed using RNA isolated from the kidney tissue of a 2-day-old Hispanic female, who died from cerebral anoxia. Family history included congenital heart disease.
LUNGNOT12	pINCY	Library was constructed using RNA isolated from lung tissue removed from a 78-year-old Caucasian male during a segmental lung resection and regional lymph node resection. Pathology indicated fibrosis pleura was puckered, but not invaded. Pathology for the associated tumor tissue indicated an invasive pulmonary grade 3 adenocarcinoma. Patient history included cerebrovascular disease, arteriosclerotic coronary artery disease, thrombophlebitis, chronic obstructive pulmonary disease, and asthma. Family history included intracranial hematoma, cerebrovascular disease, arteriosclerotic coronary artery disease, and type I diabetes.
MPHGLPT02	PSPORT1	Library was constructed using RNA isolated from adherent mononuclear cells, which came from a pool of male and female donors. The cells were stimulated with LPS.
MUSCNOT07	pINCY	Library was constructed using RNA isolated from muscle tissue removed from the forearm of a 38-year-old Caucasian female during a soft tissue excision. Pathology for the associated tumor tissue indicated intramuscular hemangioma. Family history included breast cancer, benign hypertension, cerebrovascular disease, colon cancer, and type II diabetes.
MUSLTDR02	PCDNA2.1	This random primed library was constructed using RNA isolated from right lower thigh muscle tissue removed from a 58-year-old Caucasian male during a wide resection of the right posterior thigh. Pathology indicated no residual tumor was identified in the right posterior thigh soft tissue. Changes were consistent with a previous biopsy site. On section through the soft tissue and muscle there was a smooth cystic cavity with hemorrhage around the margin on one side. The wall of the cyst was smooth and pale-tan. Pathology for the matched tumor tissue indicated a grade II liposarcoma. Patient history included liposarcoma (right thigh), and hypercholesterolemia. Previous surgeries included resection of right thigh mass. Family history included myocardial infarction and an unspecified rare blood disease.
MYOMNOT01	PSPORT1	Library was constructed using RNA isolated from uterine myometrial tissue removed from a 43-year-old Caucasian female during a vaginal hysterectomy and removal of the fallopian tubes and ovaries. Family history included lung cancer, stroke, type II diabetes, hepatic lesion, chronic liver disease, hyperlipidemia, congenital heart anomaly, and mitral valve prolapse.

Table 6

Library	Vector	Library Description
PANCTUT02	pINCY	Library was constructed using RNA isolated from pancreatic tumor tissue removed from a 45-year-old Caucasian female during radical pancreaticoduodenectomy. Pathology indicated a grade 4 anaplastic carcinoma. Family history included benign hypertension, hyperlipidemia and atherosclerotic coronary artery disease.
PLACNOB01	PBLUESCRIPT	Library was constructed using RNA isolated from placenta.
PLACNOR01	PCDNA2.1	This random primed library was constructed using pooled cDNA from two different donors. cDNA was generated using mRNA isolated from placental tissue removed from a Caucasian fetus (donor A), who died after 16 weeks' gestation from fetal demise and hydrocephalus and from placental tissue removed from a Caucasian male fetus (donor B), who died after 18 weeks' gestation from fetal demise. Patient history for donor A included umbilical cord wrapped around the head (3 times) and the shoulders (1 time). Serology was positive for anti-CMV and remaining serologies were negative. Family history included multiple pregnancies and live births, and an abortion in the mother. Serology was negative for donor B.
PROSTUT03	PSPORT1	Library was constructed using RNA isolated from the prostate tumor tissue removed from a 67-year-old Caucasian male during radical prostatectomy and lymph node biopsy. Pathology indicated adenocarcinoma Gleason grade 3+3. Adenofibromatous hyperplasia was present. Patient history included coronary artery disease, stomach ulcer, and osteoarthritis. Family history included congestive heart.
SINTFER02	pINCY	This random primed library was constructed using RNA isolated from small intestine tissue removed from a Caucasian male fetus who died from fetal demise.
SKINDIA01	PSPORT1	This amplified library was constructed using RNA isolated from diseased skin tissue removed from 1 female and 4 males during skin biopsies. Pathologies indicated tuberculoid and lepromatious leprosy.
SPLNNOT04	pINCY	Library was constructed using RNA isolated from the spleen tissue of a 2-year-old Hispanic male, who died from cerebral anoxia. Past medical history and serologies were negative.

Table 6

Library	Vector	Library Description
STOMTDA01	PSPORT1	This amplified library was constructed using RNA made from pooled cDNA from three donors. cDNA was generated using mRNA isolated from stomach tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation (donor A): from stomach tissue removed from a 61-year-old male (donor B); and from stomach tissue removed from a 61-year-old Caucasian male (donor C) during a partial esophagectomy, proximal gastrectomy, pyloromyotomy, and regional lymph node excision. Pathology for the associated tumor tissue (donor B) indicated no viable tumor identified in the esophagogastrectomy specimen. A circumscribed ulceration was identified at the gastroesophageal junction associated with pools of mucin and plasma cells. Multiple perigastric lymph nodes and one subcarinal node showed metastatic adenocarcinoma. For donor C, pathology for the associated tumor indicated invasive grade 3 adenocarcinoma in the esophagus, extending distally to involve the gastroesophageal junction. The tumor extended through the muscularis to involve periesophageal and perigastric soft tissues. One perigastric and two periesophageal
		lymph nodes were positive for tumor. There were multiple perigastric and periesophageal tumor implants. Surgical margins negative (C). Serologies were negative (A). Donor C presented with deficiency anemia and myelodysplasia. Patient history (C) included hyperlipidemia, and tobacco and alcohol abuse in remission. Previous surgery (C) included adenotonsillectomy, rhinoplasty, vasectomy, and hemorrhoidectomy. Patient medications (C) included Epoetin, Danocrine, Berocca Plus tablets, Selenium, vitamin B6 phosphate, vitamins E & C, and beta carotene.
TESTNOT03	PBLUESCRIPT	Library was constructed using RNA isolated from testicular tissue removed from a 37-year-old Caucasian male, who died from liver disease. Patient history included cirrhosis, jaundice, and liver failure.
THP1AZS08	PSPORT1	This subtracted THP-1 promonocyte cell line library was constructed using 5.76 x 1e6 clones from a 5-aza-2'-deoxycytidine (AZ) treated THP-1 cell library. Starting RNA was made from THP-1 promonocyte cells treated for three days with 0.8 micromolar AZ. The donor had acute monocytic leukemia The hybridization probe for subtraction was derived from a similarly constructed library, made from 1 microgram of polyA RNA isolated from untreated THP-1 cells. 5.76 million clones from the AZ-treated THP-1 cell library were then subjected to two rounds of subtractive hybridization with 5 million clones from the untreated THP-1 cell library. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR (1991) 19:1954, and Bonaldo et al., Genome Research (1996) 6:791.

Table 6

Library	Vector	Library Description
THYMDIT01	pINCY	The library was constructed using RNA isolated from diseased thymus tissue removed from a 16-year-old Caucasian female during a total excision of thymus and regional lymph node excision. Pathology indicated thymic follicular hyperplasia. The right lateral thymus showed reactive lymph nodes. A single reactive lymph node was also identified at the inferior thymus margin. The patient presented with myasthenia gravis, malaise, fatigue, dysphagia, severe muscle weakness and prominent eyes. Patient history included frozen face muscles. Family history included depressive disorder, hepatitis B, myocardial infarction, atherosclerotic coronary artery disease, leukemia, multiple sclerosis, and lupus.
UCMCL5T01	PBLUESCRIPT	Library was constructed using RNA isolated from mononuclear cells obtained from the umbilical cord blood of 12 individuals. The cells were cultured for 12 days with IL-5 before RNA was obtained from the pooled lysates.
UTRSNOR01	pINCY	Library was constructed using RNA isolated from uterine endometrium tissue removed from a 29-year-old Caucasian female during a vaginal hysterectomy and cystocele repair. Pathology indicated the endometrium was secretory, and the cervix showed mild chronic cervicitis with focal squamous metaplasia. Pathology for the associated tumor tissue indicated intramural uterine leiomyoma. Patient history included hypothyroidism, pelvic floor relaxation, and paraplegia. Family history included benign hypertension, type II diabetes, and hyperlipidemia.
UTRSNOT01	PSPORT1	Library was constructed using RNA isolated from the uterine tissue of a 59-year-old female who died of a myocardial infarction. Patient history included cardiomyopathy, coronary artery disease, previous myocardial infarctions, hypercholesterolemia, hypotension, and arthritis.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value=1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value=1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART, and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM, INCY, SMART, or TIGRFAM hits:* Probability value=1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score=120 or greater; Match length=56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
78	2234223	1217044H1	SNP00007480	39	1754	A	A	G	noncoding	n/a	n/a	n/a	n/a
78	2234223	1217044H1	SNP00106942	44	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	1240756H1	SNP00007478	72	1239	A	A	G	E379	0.73	n/a	n/a	n/a
78	2234223	1326647H1	SNP00121124	205	953	G	G	C	R284	n/d	n/a	n/a	n/a
78	2234223	1326647H1	SNP00131843	35	783	C	C	T	L227	n/a	n/a	n/a	n/a
78	2234223	1354894H1	SNP00106717	202	1968	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	1359298H1	SNP00121124	183	953	G	G	C	R284	n/d	n/a	n/a	n/a
78	2234223	1359298H1	SNP00131843	13	783	C	C	T	L227	n/a	n/a	n/a	n/a
78	2234223	1363568H1	SNP00007479	168	1662	C	C	T	S520	n/a	n/a	n/a	n/a
78	2234223	1476168H1	SNP00007479	131	1662	T	C	T	S520	n/a	n/a	n/a	n/a
78	2234223	1492427H1	SNP00007479	98	1662	C	C	T	S520	n/a	n/a	n/a	n/a
78	2234223	1492427H1	SNP00007480	190	1754	A	A	G	noncoding	n/a	n/a	n/a	n/a
78	2234223	1492427H1	SNP00106942	195	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	1511450H1	SNP00121124	144	953	G	G	C	R284	n/d	n/a	n/a	n/a
78	2234223	1625360H1	SNP00007479	45	1662	C	C	T	S520	n/a	n/a	n/a	n/a
78	2234223	1625360H1	SNP00007480	137	1754	A	A	G	noncoding	n/a	n/a	n/a	n/a
78	2234223	1625360H1	SNP00106942	142	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	1629012H1	SNP00007479	53	1662	C	C	T	S520	n/a	n/a	n/a	n/a
78	2234223	1629012H1	SNP00007480	145	1754	A	A	G	noncoding	n/a	n/a	n/a	n/a
78	2234223	1629012H1	SNP00106942	150	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	1645978H1	SNP00007478	97	1239	A	A	G	E379	0.73	n/a	n/a	n/a
78	2234223	1649282H1	SNP00007478	163	1239	A	A	G	E379	0.73	n/a	n/a	n/a
78	2234223	1650057H1	SNP00007478	97	1239	A	A	G	E379	0.73	n/a	n/a	n/a
78	2234223	1666220H1	SNP00005164	93	968	T	C	T	I289	0.51	n/a	n/a	n/a
78	2234223	1712235H1	SNP00007478	171	1239	A	A	G	E379	0.73	n/a	n/a	n/a
78	2234223	1732816H1	SNP00005164	72	968	C	C	T	T289	0.51	n/a	n/a	n/a
78	2234223	1808387H1	SNP00131843	159	783	C	C	T	L227	n/a	n/a	n/a	n/a
78	2234223	1809387H1	SNP00007479	58	1662	C	C	T	S520	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
78	2234223	1809387H1	SNP00007480	150	1754	A	A	G	noncoding	n/a	n/a	n/a	n/a
78	2234223	1809387H1	SNP00106942	155	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	1921608H1	SNP00106717	66	1968	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	2005818H1	SNP00007480	29	1754	A	A	G	noncoding	n/a	n/a	n/a	n/a
78	2234223	2005818H1	SNP00106942	34	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	2118555H1	SNP00106717	195	1968	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	2234223H1	SNP00120610	45	45	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	2234223H1	SNP00120611	192	192	G	A	G	R30	n/a	n/a	n/a	n/a
78	2234223	2253789H1	SNP00007478	64	1239	A	A	G	E379	0.73	n/a	n/a	n/a
78	2234223	2302939H1	SNP00106717	150	1968	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	2593062H1	SNP00007479	240	1662	C	C	T	S520	n/a	n/a	n/a	n/a
78	2234223	2597285H1	SNP00120610	80	45	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	2615842H1	SNP00007478	64	1239	A	A	G	E379	0.73	n/a	n/a	n/a
78	2234223	2617041H2	SNP00007479	48	1662	C	C	T	S520	n/a	n/a	n/a	n/a
78	2234223	2617041H2	SNP00007480	140	1754	A	A	G	noncoding	n/a	n/a	n/a	n/a
78	2234223	2617041H2	SNP00106942	145	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	2650315H1	SNP00007478	105	1239	A	A	G	E379	0.73	n/a	n/a	n/a
78	2234223	2692439H1	SNP00065980	207	1155	T	T	C	D351	n/a	n/a	n/a	n/a
78	2234223	2805067H1	SNP00007478	146	1239	G	A	G	E379	0.73	n/a	n/a	n/a
78	2234223	2849386H1	SNP00005164	46	968	C	C	T	T289	0.51	n/a	n/a	n/a
78	2234223	2853718H1	SNP00120610	80	45	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	2902853H1	SNP00007479	111	1662	C	C	T	S520	n/a	n/a	n/a	n/a
78	2234223	2902853H1	SNP00007480	203	1754	A	A	G	noncoding	n/a	n/a	n/a	n/a
78	2234223	2902853H1	SNP00106942	208	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	2998685H1	SNP00120610	125	45	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	3024806H1	SNP00106717	15	1968	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	311059H1	SNP00007480	64	1754	A	A	G	noncoding	n/a	n/a	n/a	n/a
78	2234223	311059H1	SNP00106942	69	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
78	2234223	3147933H1	SNP00065980	81	1155	T	T	C	D351	n/a	n/a	n/a	n/a
78	2234223	3276763H1	SNP00120610	87	45	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	3341543H1	SNP00120610	78	45	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	3359026H1	SNP00120610	121	45	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	3407137H1	SNP00106717	161	1968	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	3410051H1	SNP00106717	199	1968	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	3410130H1	SNP00007480	69	1754	A	A	G	noncoding	n/a	n/a	n/a	n/a
78	2234223	3410130H1	SNP00106942	74	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	3474387H1	SNP00120610	114	45	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	3475886H1	SNP00106717	176	1968	A	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	3519074H1	SNP00005164	47	968	C	C	T	T289	0.51	n/a	n/a	n/a
78	2234223	3642160H1	SNP00121124	238	953	G	G	C	R284	n/d	n/a	n/a	n/a
78	2234223	3966605H1	SNP00007479	238	1662	C	C	T	S520	n/a	n/a	n/a	n/a
78	2234223	4042390H1	SNP00120610	137	45	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	4044876H1	SNP00007479	232	1662	C	C	T	S520	n/a	n/a	n/a	n/a
78	2234223	4140677H1	SNP00120610	109	45	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	4300939H1	SNP00014621	114	731	A	G	A	H210	n/a	n/a	n/a	n/a
78	2234223	4301779H1	SNP00027227	107	1139	T	C	T	V346	0.09	n/a	n/a	n/a
78	2234223	4382742H1	SNP00120610	106	45	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	4404495H1	SNP00106717	173	1968	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	4406258H1	SNP00131843	224	783	C	C	T	L227	n/a	n/a	n/a	n/a
78	2234223	4447374H1	SNP00106942	262	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	4491017H1	SNP00007479	87	1662	C	C	T	S520	n/a	n/a	n/a	n/a
78	2234223	4491017H1	SNP00007480	179	1754	G	A	G	noncoding	n/a	n/a	n/a	n/a
78	2234223	4491017H1	SNP00106717	393	1968	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	4491017H1	SNP00106942	184	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	4594868H1	SNP00007478	40	1239	A	A	G	E379	0.73	n/a	n/a	n/a
78	2234223	4610210H1	SNP00007479	155	1662	C	C	T	S520	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
78	2234223	4610210H1	SNP00007480	247	1754	A	A	G	noncoding	n/a	n/a	n/a	n/a
78	2234223	4610210H1	SNP00106942	252	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	4640458H1	SNP00007478	101	1239	A	A	G	E379	0.73	n/a	n/a	n/a
78	2234223	4644386H1	SNP00005164	4	968	C	C	T	T289	0.51	n/a	n/a	n/a
78	2234223	4649341H1	SNP00131843	272	783	C	C	T	L227	n/a	n/a	n/a	n/a
78	2234223	4649635H1	SNP00027227	195	1139	T	C	T	V346	0.09	n/a	n/a	n/a
78	2234223	4649635H1	SNP00121124	9	953	G	G	C	R284	n/d	n/a	n/a	n/a
78	2234223	4668201H1	SNP00007480	171	1754	G	A	G	noncoding	n/a	n/a	n/a	n/a
78	2234223	4668201H1	SNP00106942	176	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	4729718H1	SNP00007478	108	1239	A	A	G	E379	0.73	n/a	n/a	n/a
78	2234223	4731671H1	SNP00120610	126	45	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	4823623H1	SNP00027227	216	1139	C	C	T	A346	0.09	n/a	n/a	n/a
78	2234223	4823623H1	SNP00121124	31	953	G	G	C	R284	n/d	n/a	n/a	n/a
78	2234223	4823804H1	SNP00007479	218	1662	C	C	T	S520	n/a	n/a	n/a	n/a
78	2234223	4856741H1	SNP00007479	41	1662	C	C	T	S520	n/a	n/a	n/a	n/a
78	2234223	4856741H1	SNP00007480	133	1754	A	A	G	noncoding	n/a	n/a	n/a	n/a
78	2234223	4856741H1	SNP00106942	138	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	4859634H1	SNP00005164	58	968	T	C	T	I289	0.51	n/a	n/a	n/a
78	2234223	5083203H1	SNP00007478	181	1239	A	A	G	E379	0.73	n/a	n/a	n/a
78	2234223	5092057H1	SNP00121124	247	953	C	G	C	T284	n/d	n/a	n/a	n/a
78	2234223	5092057H1	SNP00131843	76	783	C	C	T	L227	n/a	n/a	n/a	n/a
78	2234223	5092153H1	SNP00121124	246	953	C	G	C	T284	n/d	n/a	n/a	n/a
78	2234223	5092153H1	SNP00131843	76	783	C	C	T	L227	n/a	n/a	n/a	n/a
78	2234223	5108452H1	SNP00007480	223	1754	A	A	G	noncoding	n/a	n/a	n/a	n/a
78	2234223	5108452H1	SNP00106942	228	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	5108794H1	SNP00005164	108	968	C	C	T	T289	0.51	n/a	n/a	n/a
78	2234223	5110168H1	SNP00007479	88	1662	C	C	T	S520	n/a	n/a	n/a	n/a
78	2234223	5110168H1	SNP00007480	180	1754	A	A	G	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
78	2234223	5110168H1	SNP00106942	185	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	5177879H1	SNP00131843	73	783	C	C	T	L227	n/a	n/a	n/a	n/a
78	2234223	5762279H1	SNP00131843	301	783	C	C	T	L227	n/a	n/a	n/a	n/a
78	2234223	5768858H1	SNP00007480	73	1754	A	A	G	noncoding	n/a	n/a	n/a	n/a
78	2234223	5768858H1	SNP00106717	287	1968	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	5768858H1	SNP00106942	78	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	5771594H1	SNP00027227	299	1139	C	C	T	A346	0.09	n/a	n/a	n/a
78	2234223	5771594H1	SNP00121124	113	953	G	G	C	R284	n/d	n/a	n/a	n/a
78	2234223	5873136H1	SNP00007479	278	1662	C	C	T	S520	n/a	n/a	n/a	n/a
78	2234223	602233H1	SNP00065980	66	1155	C	T	C	D351	n/a	n/a	n/a	n/a
78	2234223	6400621H1	SNP00131843	168	783	C	C	T	L227	n/a	n/a	n/a	n/a
78	2234223	6432481H1	SNP00005164	139	968	T	C	T	I289	0.51	n/a	n/a	n/a
78	2234223	6434272H1	SNP00007480	25	1754	A	A	G	noncoding	n/a	n/a	n/a	n/a
78	2234223	6434272H1	SNP00106942	30	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	6455080H1	SNP00121124	327	953	G	G	C	R284	n/d	n/a	n/a	n/a
78	2234223	6455080H1	SNP00131843	157	783	C	C	T	L227	n/a	n/a	n/a	n/a
78	2234223	6481329H1	SNP00065980	400	1155	T	T	C	D351	n/a	n/a	n/a	n/a
78	2234223	6487236H1	SNP00120610	79	45	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	6487236H1	SNP00120611	226	192	G	A	G	R30	n/a	n/a	n/a	n/a
78	2234223	6498901H1	SNP00007478	337	1239	A	A	G	E379	0.73	n/a	n/a	n/a
78	2234223	6499937H1	SNP00007479	274	1662	C	C	T	S520	n/a	n/a	n/a	n/a
78	2234223	6499937H1	SNP00007480	366	1754	A	A	G	noncoding	n/a	n/a	n/a	n/a
78	2234223	6499937H1	SNP00106717	580	1968	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	6499937H1	SNP00106942	371	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	6527386H1	SNP00145864	520	1755	T	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	6599352H1	SNP00007478	123	1239	A	A	G	E379	0.73	n/a	n/a	n/a
78	2234223	6874973H1	SNP00145864	429	1755	T	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	6931365H1	SNP00120610	137	45	G	G	A	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
78	2234223	6931365H1	SNP00120611	284	192	G	A	G	R30	n/a	n/a	n/a	n/a
78	2234223	7128075H1	SNP00120610	79	45	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	7128075H1	SNP00120611	226	192	G	A	G	R30	n/a	n/a	n/a	n/a
78	2234223	7171458H1	SNP00027227	388	1139	C	C	T	A346	0.09	n/a	n/a	n/a
78	2234223	7171458H1	SNP00121124	202	953	G	G	C	R284	n/d	n/a	n/a	n/a
78	2234223	7171458H1	SNP00131843	32	783	T	C	T	L227	n/a	n/a	n/a	n/a
78	2234223	7199375H1	SNP00007479	293	1662	C	C	T	S520	n/a	n/a	n/a	n/a
78	2234223	7199375H1	SNP00007480	201	1754	A	A	G	noncoding	n/a	n/a	n/a	n/a
78	2234223	7199375H1	SNP00106942	196	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	7250830H1	SNP00121124	471	953	G	G	C	R284	n/d	n/a	n/a	n/a
78	2234223	7250830H1	SNP00131843	301	783	C	C	T	L227	n/a	n/a	n/a	n/a
78	2234223	7264645H1	SNP00121124	471	953	G	G	C	R284	n/d	n/a	n/a	n/a
78	2234223	7264645H1	SNP00131843	301	783	C	C	T	L227	n/a	n/a	n/a	n/a
78	2234223	7356568H1	SNP00007480	77	1754	A	A	G	noncoding	n/a	n/a	n/a	n/a
78	2234223	7356568H1	SNP00106717	291	1968	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	7356568H1	SNP00106942	82	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	7434736H1	SNP00007478	75	1239	A	A	G	E379	0.73	n/a	n/a	n/a
78	2234223	7634724J1	SNP00121124	427	953	G	G	C	R284	n/d	n/a	n/a	n/a
78	2234223	7634724J1	SNP00131843	253	783	C	C	T	L227	n/a	n/a	n/a	n/a
78	2234223	7639136H1	SNP00062174	519	560	T	T	C	L153	n/a	n/a	n/a	n/a
78	2234223	7639136H1	SNP00120611	150	192	G	A	G	R30	n/a	n/a	n/a	n/a
78	2234223	767074H1	SNP00106717	97	1968	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	7689370H1	SNP00121124	409	953	G	G	C	R284	n/d	n/a	n/a	n/a
78	2234223	7689370H1	SNP00131843	239	783	C	C	T	L227	n/a	n/a	n/a	n/a
78	2234223	7729027H1	SNP00120610	138	45	A	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	7729027H1	SNP00120611	285	192	G	A	G	R30	n/a	n/a	n/a	n/a
78	2234223	7754160H1	SNP00062174	580	560	T	T	C	L153	n/a	n/a	n/a	n/a
78	2234223	7754160H1	SNP00120610	67	45	G	G	A	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
78	2234223	7754160H1	SNP00120611	213	192	G	A	G	R30	n/a	n/a	n/a	n/a
78	2234223	7754160J1	SNP00007479	322	1662	C	C	T	S520	n/a	n/a	n/a	n/a
78	2234223	7754160J1	SNP00007480	230	1754	A	A	G	noncoding	n/a	n/a	n/a	n/a
78	2234223	7754160J1	SNP00106717	16	1968	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	7754160J1	SNP00106942	225	1759	T	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	838792H1	SNP00145864	211	1755	C	C	T	noncoding	n/a	n/a	n/a	n/a
78	2234223	8513326H1	SNP00131843	421	783	C	C	T	L227	n/a	n/a	n/a	n/a
78	2234223	898332H1	SNP00131843	190	783	C	C	T	L227	n/a	n/a	n/a	n/a
78	2234223	899610H1	SNP00120610	131	45	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	909095H1	SNP00106717	67	1968	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	2234223	915655H1	SNP00007478	205	1239	A	A	G	E379	0.73	n/a	n/a	n/a
78	2234223	917012H1	SNP00007478	205	1239	A	A	G	E379	0.73	n/a	n/a	n/a

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-39,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:2-5, SEQ ID NO:7, SEQ ID NO:10-13, SEQ ID NO:15-20, SEQ ID NO:22-32, SEQ ID NO:34-36 and SEQ ID NO:38,

c) a polypeptide comprising a naturally occurring amino acid sequence at least 92% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:14 and SEQ ID NO:33,

d) a polypeptide comprising a naturally occurring amino acid sequence at least 96% identical to the amino acid sequence of SEQ ID NO:37,

e) a polypeptide consisting essentially of a naturally occurring amino acid sequence at least 90% identical to the amino acid sequence of SEQ ID NO:39,

f) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, and

g) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-39.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-39.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:40-78.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-39.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:40-78,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:42-46, SEQ ID NO:48-59, SEQ ID NO:61-71, SEQ ID NO:73-75 and SEQ ID NO:77,

c) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 92% identical to the polynucleotide sequence of SEQ ID NO:72,

d) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 96% identical to the polynucleotide sequence of SEQ ID NO:76,

e) a polynucleotide consisting essentially of a naturally occurring polynucleotide sequence at least 90% identical to the polynucleotide sequence of SEQ ID NO:78,

f) a polynucleotide complementary to a polynucleotide of a),

g) a polynucleotide complementary to a polynucleotide of b),

h) a polynucleotide complementary to a polynucleotide of c),

i) a polynucleotide complementary to a polynucleotide of d),

j) a polynucleotide complementary to a polynucleotide of e), and

k) an RNA equivalent of a)-j).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-39.

19. A method for treating a disease or condition associated with decreased expression of functional MDDT, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional MDDT, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional MDDT, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of MDDT in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable

for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of MDDT in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of MDDT in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibodies from said animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-39.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-39.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-39 in a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of

SEQ ID NO:1-39 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-39 from a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-39.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-39, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-39.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-39 in a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of

SEQ ID NO:1-39 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-39 from a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-39.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is

completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.

76. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.

77. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:22.

78. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:23.

79. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:24.

80. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:25.

81. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:26.

82. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:27.

83. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:28.

84. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:29.

85. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:30.

86. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:31.

87. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:32.

88. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:33.

89. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:34.

90. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:35.

91. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:36.

92. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:37.

93. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:38.

94. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:39.

95. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:40.

96. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:41.

97. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:42.

98. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:43.

99. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:44.

100. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:45.

101. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:46.

102. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:47.

103. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:48.

104. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:49.

105. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:50.

106. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:51.

107. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:52.

108. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:53.

109. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:54.

110. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID

NO:55.

111. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:56.

112. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:57.

113. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:58.

114. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:59.

115. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:60.

116. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:61.

117. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:62.

118. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:63.

119. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:64.

120. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:65.

121. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID

<110> INCYTE GENOMICS, INC.
TANG, Y. TOM
YUE, Henry
BAUGHN, Mariah R.
DUGGAN, Brendan M.
WARREN, Bridget A.
BANDMAN, Olga
RICHARDSON, Thomas W.
BURFORD, Neil
SANJANWALA, Bharati
BECHA, Shanya D.
YAO, Monique G.
YANG, Junming
TRAN, Uyen K.
HAFALIA, April J.A.
GRIFFIN, Jennifer A.
SWARNAKER, Anita
ELLIOTT, Vicki S.
RECIPON, Shirley A.
KHAN, Farrah A.
LEE, Ernestine A.
YUE, Huibin
LU, Dyung Aina M.
WALIA, Narinder K.
THANGAVELU, Kavitha
ARVIZU, Chandra S.
XU, Yuming
ISON, Craig H.
HUANG, Jiaqi
DING, Li
HONCHELL, Cynthia D.
BOROWSKY, Mark L.
EMERLING, Brooke M.
PETERSON, David P.
LU, Yan
RAMKUMAR, Jayalaxmi
MASON, Patricia M.
ZEBARJADIAN, Yeganeh
AZIMZAI, Yalda
STUVE, Laura L.
KAMIGAKI, Laura Y.
BARROSO, Ines
LEE, Sally
KABLE, Amy E.

<120> MOLECULES FOR DISEASE DETECTION AND TREATMENT

<130> PF-0991 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/293,723; 60/295,257; 60/297,220; 60/300,526; 60/301,874;
60/359,413
<151> 2001-05-25; 2001-06-01; 2001-06-08; 2001-06-21; 2001-06-29;
2002-02-22

<160> 78
<170> PERL Program

<210> 1
<211> 200
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4973222CD1

<400> 1
Met Asn Ser Phe Ser Thr Ser Ala Phe Gly Pro Val Ala Phe Pro
1 5 10 15
Ala Pro Val Pro Pro Gly Glu Asp Ser Lys Asp Val Ala Ala Pro
20 25 30
His Arg Gln Pro Leu Thr Ser Ser Glu Arg Ile Asp Lys Gln Ile
35 40 45
Arg His Ile Leu Asp Gly Ile Ser Ala Leu Arg Lys Glu Thr Cys
50 55 60
Asn Lys Ser Asn Met Cys Glu Ser Ser Lys Glu Ala Leu Ala Glu
65 70 75
Asn Asn Leu Asn Leu Pro Lys Met Ala Glu Lys Asp Gly Cys Phe
80 85 90
Gln Ser Gly Phe Asn Glu Glu Thr Cys Leu Val Lys Ile Ile Thr
95 100 105
Gly Leu Leu Glu Phe Glu Val Tyr Leu Glu Tyr Leu Gln Asn Arg
110 115 120
Phe Glu Ser Ser Glu Glu Gln Ala Arg Ala Val Gln Met Ser Thr
125 130 135
Lys Val Leu Val Gln Phe Leu Gln Lys Lys Ala Lys Asn Leu Asp
140 145 150
Ala Ile Thr Thr Pro Asp Pro Thr Thr Asn Ala Ser Leu Leu Thr
155 160 165
Lys Leu Gln Ala Gln Asn Gln Trp Leu Gln Asp Met Thr Thr Arg
170 175 180
Leu Ile Leu Arg Ser Phe Lys Glu Phe Leu Gln Ser Ser Leu Arg
185 190 195
Ala Leu Arg Gln Met
200

<210> 2
<211> 663
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 55009060CD1

<400> 2
Met Ala Gly Met Lys Thr Ala Ser Gly Asp Tyr Ile Asp Ser Ser
1 5 10 15
Trp Glu Leu Arg Val Phe Val Gly Glu Glu Asp Pro Glu Ala Glu
20 25 30

```
Ser Val Thr Leu Arg Val Thr Gly Glu Ser His Ile Gly Gly Val
                35                  40                  45
Leu Leu Lys Ile Val Glu Gln Ile Asn Arg Lys Gln Asp Trp Ser
                50                  55                  60
Asp His Ala Ile Trp Trp Glu Gln Lys Arg Gln Trp Leu Leu Gln
                65                  70                  75
Thr His Trp Thr Leu Asp Lys Tyr Gly Ile Leu Ala Asp Ala Arg
                80                  85                  90
Leu Phe Phe Gly Pro Gln His Arg Pro Val Ile Leu Arg Leu Pro
                95                 100                 105
Asn Arg Arg Ala Leu Arg Leu Arg Ala Ser Phe Ser Gln Pro Leu
               110                 115                 120
Phe Gln Ala Val Ala Ala Ile Cys Arg Leu Leu Ser Ile Arg His
               125                 130                 135
Pro Glu Glu Leu Ser Leu Leu Arg Ala Pro Glu Lys Lys Glu Lys
               140                 145                 150
Lys Lys Lys Glu Lys Glu Pro Glu Glu Glu Leu Tyr Asp Leu Ser
               155                 160                 165
Lys Val Val Leu Ala Gly Gly Val Ala Pro Ala Leu Phe Arg Gly
               170                 175                 180
Met Pro Ala His Phe Ser Asp Ser Ala Gln Thr Glu Ala Cys Tyr
               185                 190                 195
His Met Leu Ser Arg Pro Gln Pro Pro Pro Asp Pro Leu Leu Leu
               200                 205                 210
Gln Arg Leu Pro Arg Pro Ser Ser Leu Ser Asp Lys Thr Gln Leu
               215                 220                 225
His Ser Arg Trp Leu Asp Ser Ser Arg Cys Leu Met Gln Gln Gly
               230                 235                 240
Ile Lys Ala Gly Asp Ala Leu Trp Leu Arg Phe Lys Tyr Tyr Ser
               245                 250                 255
Phe Phe Asp Leu Asp Pro Lys Thr Asp Pro Val Arg Leu Thr Gln
               260                 265                 270
Leu Tyr Glu Gln Ala Arg Trp Asp Leu Leu Leu Glu Glu Ile Asp
               275                 280                 285
Cys Thr Glu Glu Glu Met Met Val Phe Ala Ala Leu Gln Tyr His
               290                 295                 300
Ile Asn Lys Leu Ser Gln Ser Gly Glu Val Gly Glu Pro Ala Gly
               305                 310                 315
Thr Asp Pro Gly Leu Asp Asp Leu Asp Val Ala Leu Ser Asn Leu
               320                 325                 330
Glu Val Lys Leu Glu Gly Ser Ala Pro Thr Asp Val Leu Asp Ser
               335                 340                 345
Leu Thr Thr Ile Pro Glu Leu Lys Asp His Leu Arg Ile Phe Arg
               350                 355                 360
Pro Arg Lys Leu Thr Leu Lys Gly Tyr Arg Gln His Trp Val Val
               365                 370                 375
Phe Lys Glu Thr Thr Leu Ser Tyr Tyr Lys Ser Gln Asp Glu Ala
               380                 385                 390
Pro Gly Asp Pro Ile Gln Gln Leu Asn Leu Lys Gly Cys Glu Val
               395                 400                 405
Val Pro Asp Val Asn Val Ser Gly Gln Lys Phe Cys Ile Lys Leu
               410                 415                 420
Leu Val Pro Ser Pro Glu Gly Met Ser Glu Ile Tyr Leu Arg Cys
               425                 430                 435
Gln Asp Glu Gln Gln Tyr Ala Arg Trp Met Ala Gly Cys Arg Leu
               440                 445                 450
```

```
Ala Ser Lys Gly Arg Thr Met Ala Asp Ser Ser Tyr Thr Ser Glu
                455                 460                 465
Val Gln Ala Ile Leu Ala Phe Leu Ser Leu Gln Arg Thr Gly Ser
                470                 475                 480
Gly Gly Pro Gly Asn His Pro His Gly Pro Asp Ala Ser Ala Glu
                485                 490                 495
Gly Leu Asn Pro Tyr Gly Leu Val Ala Pro Arg Phe Gln Arg Lys
                500                 505                 510
Phe Lys Ala Lys Gln Leu Thr Pro Arg Ile Leu Glu Ala His Gln
                515                 520                 525
Asn Val Ala Gln Leu Ser Leu Ala Glu Ala Gln Leu Arg Phe Ile
                530                 535                 540
Gln Ala Trp Gln Ser Leu Pro Asp Phe Gly Ile Ser Tyr Val Met
                545                 550                 555
Val Arg Phe Lys Gly Ser Arg Lys Asp Glu Ile Leu Gly Ile Ala
                560                 565                 570
Asn Asn Arg Leu Ile Arg Ile Asp Leu Ala Val Gly Asp Val Val
                575                 580                 585
Lys Thr Trp Arg Phe Ser Asn Met Arg Gln Trp Asn Val Asn Trp
                590                 595                 600
Asp Ile Arg Gln Val Ala Ile Glu Phe Asp Glu His Ile Asn Val
                605                 610                 615
Ala Phe Ser Cys Val Ser Ala Ser Cys Arg Ile Val His Glu Tyr
                620                 625                 630
Ile Gly Gly Tyr Ile Phe Leu Ser Thr Arg Glu Arg Ala Arg Gly
                635                 640                 645
Glu Glu Leu Asp Glu Asp Leu Phe Leu Gln Leu Thr Gly Gly His
                650                 655                 660
Glu Ala Phe
```

<210> 3
<211> 219
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1985092CD1

<400> 3

```
Met Ser Glu Gln Glu Ala Gln Ala Pro Gly Gly Arg Gly Leu Pro
  1               5                  10                  15
Pro Asp Met Leu Ala Glu Gln Val Glu Leu Trp Trp Ser Gln Gln
                 20                  25                  30
Pro Arg Arg Ser Ala Leu Cys Phe Val Val Ala Val Gly Leu Val
                 35                  40                  45
Ala Gly Cys Gly Ala Gly Gly Val Ala Leu Leu Ser Thr Thr Ser
                 50                  55                  60
Ser Arg Ser Gly Glu Trp Arg Leu Ala Thr Gly Thr Val Leu Cys
                 65                  70                  75
Leu Leu Ala Leu Leu Val Leu Val Lys Gln Leu Met Ser Ser Ala
                 80                  85                  90
Val Gln Asp Met Asn Cys Ile Arg Gln Ala His His Val Ala Leu
                 95                 100                 105
Leu Arg Ser Gly Gly Gly Ala Asp Ala Leu Val Val Leu Leu Ser
```

```
                110                 115                 120
Gly Leu Val Leu Leu Val Thr Gly Leu Thr Leu Ala Gly Leu Ala
                125                 130                 135
Ala Ala Pro Ala Pro Ala Arg Pro Leu Ala Ala Met Leu Ser Val
                140                 145                 150
Gly Ile Ala Leu Ala Ala Leu Gly Ser Leu Leu Leu Leu Gly Leu
                155                 160                 165
Leu Leu Tyr Gln Val Gly Val Ser Gly His Cys Pro Ser Ile Cys
                170                 175                 180
Met Ala Thr Pro Ser Thr His Ser Gly His Gly Gly His Gly Ser
                185                 190                 195
Ile Phe Ser Ile Ser Gly Gln Leu Ser Ala Gly Arg Arg His Glu
                200                 205                 210
Thr Thr Ser Ser Ile Ala Ser Leu Ile
                215
```

<210> 4
<211> 318
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1553593CD1

<400> 4

```
Met Asn Thr Arg Asn Arg Val Val Asn Ser Gly Leu Gly Ala Ser
  1               5                  10                  15
Pro Ala Ser Arg Pro Thr Arg Asp Pro Gln Asp Pro Ser Gly Arg
                20                  25                  30
Gln Gly Glu Leu Ser Pro Val Glu Asp Gln Arg Glu Gly Leu Glu
                35                  40                  45
Ala Ala Pro Lys Gly Pro Ser Arg Glu Ser Val Val His Ala Gly
                50                  55                  60
Gln Arg Arg Thr Ser Ala Tyr Thr Leu Ile Ala Pro Asn Ile Asn
                65                  70                  75
Arg Arg Asn Glu Ile Gln Arg Ile Ala Glu Gln Glu Leu Ala Asn
                80                  85                  90
Leu Glu Lys Trp Lys Glu Gln Asn Arg Ala Lys Pro Val His Leu
                95                 100                 105
Val Pro Arg Arg Leu Gly Gly Ser Gln Ser Glu Thr Glu Val Arg
                110                115                 120
Gln Lys Gln Gln Leu Gln Leu Met Gln Ser Lys Tyr Lys Gln Lys
                125                130                 135
Leu Lys Arg Glu Glu Ser Val Arg Ile Lys Lys Glu Ala Glu Glu
                140                145                 150
Ala Glu Leu Gln Lys Met Lys Ala Ile Gln Arg Glu Lys Ser Asn
                155                160                 165
Lys Leu Glu Glu Lys Lys Arg Leu Gln Glu Asn Leu Arg Arg Glu
                170                175                 180
Ala Phe Arg Glu His Gln Gln Tyr Lys Thr Ala Glu Phe Leu Ser
                185                190                 195
Lys Leu Asn Thr Glu Ser Pro Asp Arg Ser Ala Cys Gln Ser Ala
                200                205                 210
Val Cys Gly Pro Gln Ser Ser Thr Trp Lys Leu Pro Ile Leu Pro
                215                220                 225
```

```
Arg Asp His Ser Trp Ala Arg Ser Trp Ala Tyr Arg Asp Ser Leu
                230                 235                 240
Lys Ala Glu Glu Asn Arg Lys Leu Gln Lys Met Lys Asp Glu Gln
                245                 250                 255
His Gln Lys Ser Glu Leu Leu Glu Leu Lys Arg Gln Gln Gln Glu
                260                 265                 270
Gln Glu Arg Ala Lys Ile His Gln Thr Glu His Arg Arg Val Asn
                275                 280                 285
Asn Ala Phe Leu Asp Arg Leu Gln Gly Lys Ser Gln Pro Gly Gly
                290                 295                 300
Leu Glu Gln Ser Gly Gly Cys Trp Asn Met Asn Ser Gly Asn Ser
                305                 310                 315
Trp Gly Ile
```

<210> 5
<211> 387
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1954122CD1

<400> 5

```
Met Glu Lys Ile Glu Glu Gln Phe Ala Asn Leu His Ile Val Lys
  1               5                  10                  15
Cys Ser Leu Gly Thr Lys Glu Pro Thr Tyr Leu Leu Gly Ile Asp
                20                  25                  30
Thr Ser Lys Thr Val Gln Ala Gly Lys Glu Asn Leu Val Ala Val
                35                  40                  45
Leu Cys Ser Asn Gly Ser Ile Arg Ile Tyr Asp Lys Glu Arg Leu
                50                  55                  60
Asn Val Leu Arg Glu Phe Ser Gly Tyr Pro Gly Leu Leu Asn Gly
                65                  70                  75
Val Arg Phe Ala Asn Ser Cys Asp Ser Val Tyr Ser Ala Cys Thr
                80                  85                  90
Asp Gly Thr Val Lys Cys Trp Asp Ala Arg Val Ala Arg Glu Lys
                95                 100                 105
Pro Val Gln Leu Phe Lys Gly Tyr Pro Ser Asn Ile Phe Ile Ser
                110                 115                 120
Phe Asp Ile Asn Cys Asn Asp His Ile Ile Cys Ala Gly Thr Glu
                125                 130                 135
Lys Val Asp Asp Asp Ala Leu Leu Val Phe Trp Asp Ala Arg Met
                140                 145                 150
Asn Ser Gln Asn Leu Ser Thr Thr Lys Asp Ser Leu Gly Ala Tyr
                155                 160                 165
Ser Glu Thr His Ser Asp Asp Val Thr Gln Val Arg Phe His Pro
                170                 175                 180
Ser Asn Pro Asn Met Val Val Ser Gly Ser Ser Asp Gly Leu Val
                185                 190                 195
Asn Val Phe Asp Ile Asn Ile Asp Asn Glu Glu Asp Ala Leu Val
                200                 205                 210
Thr Thr Cys Asn Ser Ile Ser Ser Val Ser Cys Ile Gly Trp Ser
                215                 220                 225
Gly Lys Gly Tyr Lys Gln Ile Tyr Cys Met Thr His Asp Glu Gly
```

```
                230                 235                 240
Phe Tyr Trp Trp Asp Leu Asn His Leu Asp Thr Asp Glu Pro Val
                245                 250                 255
Thr Arg Leu Asn Ile Gln Asp Val Arg Glu Val Val Asn Met Lys
                260                 265                 270
Glu Asp Ala Leu Asp Tyr Leu Ile Gly Gly Leu Tyr His Glu Lys
                275                 280                 285
Thr Asp Thr Leu His Val Ile Gly Gly Thr Asn Lys Gly Arg Ile
                290                 295                 300
His Leu Met Asn Cys Ser Met Ser Gly Leu Thr His Val Thr Ser
                305                 310                 315
Leu Gln Gly Gly His Ala Ala Thr Val Arg Ser Phe Cys Trp Asn
                320                 325                 330
Val Gln Asp Asp Ser Leu Leu Thr Gly Gly Glu Asp Ala Gln Leu
                335                 340                 345
Leu Leu Trp Lys Pro Gly Ala Ile Glu Lys Thr Phe Thr Lys Lys
                350                 355                 360
Glu Ser Met Lys Ile Ala Ser Ser Val His Gln Arg Val Arg Val
                365                 370                 375
His Ser Asn Asp Ser Tyr Lys Arg Arg Lys Lys Gln
                380                 385
```

<210> 6
<211> 577
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3159276CD1

<400> 6

```
Met Met Ser Leu Ser Pro Ile Ser Leu Gly Ala Leu Glu Gly Arg
  1               5                  10                  15
Asn Val Asp Leu Leu Gln Ile Ser Gly Gln Leu Ser Pro Arg Leu
                 20                  25                  30
Phe Arg Lys Leu Pro Pro Arg Val Cys Val Ser Leu Lys Asn Ile
                 35                  40                  45
Val Asp Glu Asp Phe Leu Tyr Ala Gly His Ile Phe Leu Gly Phe
                 50                  55                  60
Ser Lys Cys Gly Arg Tyr Val Leu Ser Tyr Thr Ser Ser Ser Gly
                 65                  70                  75
Asp Asp Asp Phe Ser Phe Tyr Ile Tyr His Leu Tyr Trp Trp Glu
                 80                  85                  90
Phe Asn Val His Ser Lys Leu Lys Leu Val Arg Gln Val Arg Leu
                 95                 100                 105
Phe Gln Asp Glu Glu Ile Tyr Ser Asp Leu Tyr Leu Thr Val Cys
                110                 115                 120
Glu Trp Pro Ser Asp Ala Ser Lys Val Ile Val Phe Gly Phe Asn
                125                 130                 135
Thr Arg Ser Ala Asn Gly Met Leu Met Asn Met Met Met Met Ser
                140                 145                 150
Asp Glu Asn His Arg Asp Ile Tyr Val Ser Thr Val Ala Val Pro
                155                 160                 165
Pro Pro Gly Arg Cys Ala Ala Cys Gln Asp Ala Ser Arg Ala His
                170                 175                 180
```

```
Pro Gly Asp Pro Asn Ala Gln Cys Leu Arg His Gly Phe Met Leu
                185                 190                 195
His Thr Lys Tyr Gln Val Val Tyr Pro Phe Pro Thr Phe Gln Pro
                200                 205                 210
Ala Phe Gln Leu Lys Lys Asp Gln Val Val Leu Leu Asn Thr Ser
                215                 220                 225
Tyr Ser Leu Val Ala Cys Ala Val Ser Val His Ser Ala Gly Asp
                230                 235                 240
Arg Ser Phe Cys Gln Ile Leu Tyr Asp His Ser Thr Cys Pro Leu
                245                 250                 255
Ala Pro Ala Ser Pro Pro Glu Pro Gln Ser Pro Glu Leu Pro Pro
                260                 265                 270
Ala Leu Pro Ser Phe Cys Pro Glu Ala Ala Pro Ala Arg Ser Ser
                275                 280                 285
Gly Ser Pro Glu Pro Ser Pro Ala Ile Ala Lys Ala Lys Glu Phe
                290                 295                 300
Val Ala Asp Ile Phe Arg Arg Ala Lys Glu Ala Lys Gly Gly Val
                305                 310                 315
Pro Glu Glu Ala Arg Pro Ala Leu Cys Pro Gly Pro Ser Gly Ser
                320                 325                 330
Arg Cys Arg Ala His Ser Glu Pro Leu Ala Leu Cys Gly Glu Thr
                335                 340                 345
Ala Pro Arg Asp Ser Pro Pro Ala Ser Glu Ala Pro Ala Ser Glu
                350                 355                 360
Pro Gly Tyr Val Asn Tyr Thr Lys Leu Tyr Tyr Val Leu Glu Ser
                365                 370                 375
Gly Glu Gly Thr Glu Pro Glu Asp Glu Leu Glu Asp Asp Lys Ile
                380                 385                 390
Ser Leu Pro Phe Val Val Thr Asp Leu Arg Gly Arg Asn Leu Arg
                395                 400                 405
Pro Met Arg Glu Arg Thr Ala Val Gln Gly Gln Tyr Leu Thr Val
                410                 415                 420
Glu Gln Leu Thr Leu Asp Phe Glu Tyr Val Ile Asn Glu Val Ile
                425                 430                 435
Arg His Asp Ala Thr Trp Gly His Gln Phe Cys Ser Phe Ser Asp
                440                 445                 450
Tyr Asp Ile Val Ile Leu Glu Val Cys Pro Glu Thr Asn Gln Val
                455                 460                 465
Leu Ile Asn Ile Gly Leu Leu Leu Leu Ala Phe Pro Ser Pro Thr
                470                 475                 480
Glu Glu Gly Gln Leu Arg Pro Lys Thr Tyr His Thr Ser Leu Lys
                485                 490                 495
Val Ala Trp Asp Leu Asn Thr Gly Ile Phe Glu Thr Val Ser Val
                500                 505                 510
Gly Asp Leu Thr Glu Val Lys Gly Gln Thr Ser Gly Ser Val Trp
                515                 520                 525
Ser Ser Tyr Arg Lys Ser Cys Val Asp Met Val Met Lys Trp Leu
                530                 535                 540
Val Pro Glu Ser Ser Gly Arg Tyr Val Asn Arg Met Thr Asn Glu
                545                 550                 555
Ala Leu His Lys Gly Cys Ser Leu Lys Val Leu Ala Asp Ser Glu
                560                 565                 570
Arg Tyr Thr Trp Ile Val Leu
                575
```

<210> 7

<211> 224
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 140052CD1

<400> 7
Met Ala Ala Pro Ser Glu Val Ala Ala Ile Ala Pro Gly Glu Gly
1 5 10 15
Asp Gly Gly Gly Gly Gly Phe Gly Ser Trp Leu Asp Gly Arg Leu
20 25 30
Glu Ala Leu Gly Val Asp Arg Ala Val Tyr Gly Ala Tyr Ile Leu
35 40 45
Gly Ile Leu Gln Glu Glu Glu Glu Glu Glu Lys Leu Asp Ala Leu
50 55 60
Gln Gly Ile Leu Ser Ala Phe Leu Glu Glu Asp Ser Leu Leu Asn
65 70 75
Ile Cys Lys Glu Ile Val Glu Arg Trp Ser Glu Thr Gln Asn Val
80 85 90
Val Thr Lys Val Lys Lys Glu Asp Glu Val Gln Ala Ile Ala Thr
95 100 105
Leu Ile Glu Lys Gln Ala Gln Ile Val Val Lys Pro Arg Met Val
110 115 120
Ser Glu Glu Glu Lys Gln Arg Lys Ala Ala Leu Leu Ala Gln Tyr
125 130 135
Ala Asp Val Thr Asp Glu Glu Asp Glu Ala Asp Glu Lys Asp Asp
140 145 150
Ser Gly Ala Thr Thr Met Asn Ile Gly Ser Asp Lys Leu Leu Phe
155 160 165
Arg Asn Thr Asn Val Glu Asp Val Leu Asn Ala Arg Lys Leu Glu
170 175 180
Arg Asp Ser Leu Arg Asp Glu Ser Gln Arg Lys Lys Glu Gln Asp
185 190 195
Lys Leu Gln Arg Glu Arg Asp Lys Leu Ala Lys Gln Glu Arg Lys
200 205 210
Glu Lys Glu Lys Lys Arg Thr Gln Arg Gly Glu Arg Lys Arg
215 220

<210> 8
<211> 600
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5158048CD1

<400> 8
Met Gly Lys Lys Leu Asp Leu Ser Lys Leu Thr Asp Glu Glu Ala
1 5 10 15
Gln His Val Leu Glu Val Val Gln Arg Asp Phe Asp Leu Arg Arg
20 25 30
Lys Glu Glu Glu Arg Leu Glu Ala Leu Lys Gly Lys Ile Lys Lys
35 40 45

```
Glu Ser Ser Lys Arg Glu Leu Leu Ser Asp Thr Ala His Leu Asn
                50                  55                  60
Glu Thr His Cys Ala Arg Cys Leu Gln Pro Tyr Gln Leu Leu Val
                65                  70                  75
Asn Ser Lys Arg Gln Cys Leu Glu Cys Gly Leu Phe Thr Cys Lys
                80                  85                  90
Ser Cys Gly Arg Val His Pro Glu Glu Gln Gly Trp Ile Cys Asp
                95                 100                 105
Pro Cys His Leu Ala Arg Val Val Lys Ile Gly Ser Leu Glu Trp
               110                 115                 120
Tyr Tyr Glu His Val Lys Ala Arg Phe Lys Arg Phe Gly Ser Ala
               125                 130                 135
Lys Val Ile Arg Ser Leu His Gly Arg Leu Gln Gly Gly Ala Gly
               140                 145                 150
Pro Glu Leu Ile Ser Glu Glu Arg Ser Gly Asp Ser Asp Gln Thr
               155                 160                 165
Asp Glu Asp Gly Glu Pro Gly Ser Glu Ala Gln Ala Gln Ala Gln
               170                 175                 180
Pro Phe Gly Ser Lys Lys Lys Arg Leu Leu Ser Val His Asp Phe
               185                 190                 195
Asp Phe Glu Gly Asp Ser Asp Asp Ser Thr Gln Pro Gln Gly His
               200                 205                 210
Ser Leu His Leu Ser Ser Val Pro Glu Ala Arg Asp Ser Pro Gln
               215                 220                 225
Ser Leu Thr Asp Glu Ser Cys Ser Glu Lys Ala Ala Pro His Lys
               230                 235                 240
Ala Glu Gly Leu Glu Glu Ala Asp Thr Gly Ala Ser Gly Cys His
               245                 250                 255
Ser His Pro Glu Glu Gln Pro Thr Ser Ile Ser Pro Ser Arg His
               260                 265                 270
Gly Ala Leu Ala Glu Leu Cys Pro Pro Gly Gly Ser His Arg Met
               275                 280                 285
Ala Leu Gly Thr Ala Ala Ala Leu Gly Ser Asn Val Ile Arg Asn
               290                 295                 300
Glu Gln Leu Pro Leu Gln Tyr Leu Ala Asp Val Asp Thr Ser Asp
               305                 310                 315
Glu Glu Ser Ile Arg Ala His Val Met Ala Ser His His Ser Lys
               320                 325                 330
Arg Arg Gly Arg Ala Ser Ser Glu Ser Gln Ile Phe Glu Leu Asn
               335                 340                 345
Lys Arg Ile Ser Ala Val Glu Cys Leu Leu Thr Tyr Leu Glu Asn
               350                 355                 360
Thr Val Val Pro Pro Leu Ala Lys Gly Leu Gly Ala Gly Val Arg
               365                 370                 375
Thr Glu Ala Asp Val Glu Glu Glu Ala Leu Arg Arg Lys Leu Glu
               380                 385                 390
Glu Leu Thr Ser Asn Val Ser Asp Gln Glu Thr Ser Ser Glu Glu
               395                 400                 405
Glu Glu Ala Lys Asp Glu Lys Ala Glu Pro Asn Arg Asp Lys Ser
               410                 415                 420
Val Gly Pro Leu Pro Gln Ala Asp Pro Glu Val Gly Thr Ala Ala
               425                 430                 435
His Gln Thr Asn Arg Gln Glu Lys Ser Pro Gln Asp Pro Gly Asp
               440                 445                 450
Pro Val Gln Tyr Asn Arg Thr Thr Asp Glu Glu Leu Ser Glu Leu
               455                 460                 465
```

```
Glu Asp Arg Val Ala Val Thr Ala Ser Glu Val Gln Gln Ala Glu
                470                 475                 480
Ser Glu Val Ser Asp Ile Glu Ser Arg Ile Ala Ala Leu Arg Ala
                485                 490                 495
Ala Gly Leu Thr Val Lys Pro Ser Gly Lys Pro Arg Arg Lys Ser
                500                 505                 510
Asn Leu Pro Ile Phe Leu Pro Arg Val Ala Gly Lys Leu Gly Lys
                515                 520                 525
Arg Pro Glu Asp Pro Asn Ala Asp Pro Ser Ser Glu Ala Lys Ala
                530                 535                 540
Met Ala Val Pro Tyr Leu Leu Arg Arg Lys Phe Ser Asn Ser Leu
                545                 550                 555
Lys Ser Gln Gly Lys Asp Asp Asp Ser Phe Asp Arg Lys Ser Val
                560                 565                 570
Tyr Arg Gly Ser Leu Thr Gln Arg Asn Pro Asn Ala Arg Lys Gly
                575                 580                 585
Met Ala Ser His Thr Phe Ala Lys Pro Val Val Ala His Gln Ser
                590                 595                 600


<210> 9
<211> 1250
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3127541CD1

<400> 9
Met Glu Gln Leu Ser Asp Glu Glu Ile Asp His Gly Ala Glu Glu
  1               5                  10                  15
Asp Ser Asp Lys Glu Asp Gln Asp Leu Asp Lys Met Phe Gly Ala
                20                  25                  30
Trp Leu Gly Glu Leu Asp Lys Leu Thr Gln Ser Leu Asp Ser Asp
                35                  40                  45
Lys Pro Met Glu Pro Val Lys Arg Ser Pro Leu Arg Gln Glu Thr
                50                  55                  60
Asn Met Ala Asn Phe Ser Tyr Arg Phe Ser Ile Tyr Asn Leu Asn
                65                  70                  75
Glu Ala Leu Asn Gln Gly Glu Thr Val Asp Leu Asp Ala Leu Met
                80                  85                  90
Ala Asp Leu Cys Ser Ile Glu Gln Glu Leu Ser Ser Ile Gly Ser
                95                 100                 105
Gly Asn Ser Lys Arg Gln Ile Thr Glu Thr Lys Ala Thr Gln Lys
               110                 115                 120
Leu Pro Val Ser Arg His Thr Leu Lys His Gly Thr Leu Lys Gly
               125                 130                 135
Leu Ser Ser Ser Ser Asn Arg Ile Ala Lys Pro Ser His Ala Ser
               140                 145                 150
Tyr Ser Leu Asp Asp Val Thr Ala Gln Leu Glu Gln Ala Ser Leu
               155                 160                 165
Ser Met Asp Glu Ala Ala Gln Gln Ser Val Leu Glu Asp Thr Lys
               170                 175                 180
Pro Leu Val Thr Asn Gln His Arg Arg Thr Ala Ser Ala Gly Thr
               185                 190                 195
```

```
Val Ser Asp Ala Glu Val His Ser Ile Ser Asn Ser Ser His Ser
                200                 205                 210
Ser Ile Thr Ser Ala Ala Ser Ser Met Asp Ser Leu Asp Ile Asp
                215                 220                 225
Lys Val Thr Arg Pro Gln Glu Leu Asp Leu Thr His Gln Gly Gln
                230                 235                 240
Pro Ile Thr Glu Glu Glu Gln Ala Ala Lys Leu Lys Ala Glu Lys
                245                 250                 255
Ile Arg Val Ala Leu Glu Lys Ile Lys Glu Ala Gln Val Lys Lys
                260                 265                 270
Leu Val Ile Arg Val His Met Ser Asp Asp Ser Ser Lys Thr Met
                275                 280                 285
Met Val Asp Glu Arg Gln Thr Val Arg Gln Val Leu Asp Asn Leu
                290                 295                 300
Met Asp Lys Ser His Cys Gly Tyr Ser Leu Asp Trp Ser Leu Val
                305                 310                 315
Glu Thr Val Ser Glu Leu Gln Met Glu Arg Ile Phe Glu Asp His
                320                 325                 330
Glu Asn Leu Val Glu Asn Leu Leu Asn Trp Thr Arg Asp Ser Gln
                335                 340                 345
Asn Lys Leu Ile Phe Met Glu Arg Ile Glu Lys Tyr Ala Leu Phe
                350                 355                 360
Lys Asn Pro Gln Asn Tyr Leu Leu Gly Lys Lys Glu Thr Ala Glu
                365                 370                 375
Met Ala Asp Arg Asn Lys Glu Val Leu Leu Glu Glu Cys Phe Cys
                380                 385                 390
Gly Ser Ser Val Thr Val Pro Glu Ile Glu Gly Val Leu Trp Leu
                395                 400                 405
Lys Asp Asp Gly Lys Lys Ser Trp Lys Lys Arg Tyr Phe Leu Leu
                410                 415                 420
Arg Ala Ser Gly Ile Tyr Tyr Val Pro Lys Gly Lys Ala Lys Val
                425                 430                 435
Ser Arg Asp Leu Val Cys Phe Leu Gln Leu Asp His Val Asn Val
                440                 445                 450
Tyr Tyr Gly Gln Asp Tyr Arg Asn Lys Tyr Lys Ala Pro Thr Asp
                455                 460                 465
Tyr Cys Leu Val Leu Lys His Pro Gln Ile Gln Lys Lys Ser Gln
                470                 475                 480
Tyr Ile Lys Tyr Leu Cys Cys Asp Asp Val Arg Thr Leu His Gln
                485                 490                 495
Trp Val Asn Gly Ile Arg Ile Ala Lys Tyr Gly Lys Gln Leu Tyr
                500                 505                 510
Met Asn Tyr Gln Glu Ala Leu Lys Arg Thr Glu Ser Ala Tyr Asp
                515                 520                 525
Trp Thr Ser Leu Ser Ser Ser Ser Ile Lys Ser Gly Ser Ser Ser
                530                 535                 540
Ser Ser Ile Pro Glu Ser Gln Ser Asn His Ser Asn Gln Ser Asp
                545                 550                 555
Ser Gly Val Ser Asp Thr Gln Pro Ala Gly His Val Arg Ser Gln
                560                 565                 570
Ser Ile Val Ser Ser Val Phe Ser Glu Ala Trp Lys Arg Gly Thr
                575                 580                 585
Gln Leu Glu Glu Ser Ser Lys Ala Arg Met Glu Ser Met Asn Arg
                590                 595                 600
Pro Tyr Thr Ser Leu Val Pro Pro Leu Ser Pro Gln Pro Lys Ile
                605                 610                 615
```

```
Val Thr Pro Tyr Thr Ala Ser Gln Pro Ser Pro Pro Leu Pro Pro
                620                 625                 630
Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro
                635                 640                 645
Pro Pro Pro Leu Pro Ser Gln Ser Ala Pro Ser Ala Gly Ser Ala
                650                 655                 660
Ala Pro Met Phe Val Lys Tyr Ser Thr Ile Thr Arg Leu Gln Asn
                665                 670                 675
Ala Ser Gln His Ser Gly Ala Leu Phe Lys Pro Pro Thr Pro Pro
                680                 685                 690
Val Met Gln Ser Gln Ser Val Lys Pro Gln Ile Leu Val Pro Pro
                695                 700                 705
Asn Gly Val Val Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Thr
                710                 715                 720
Pro Gly Ser Ala Met Ala Gln Leu Lys Pro Ala Pro Cys Ala Pro
                725                 730                 735
Ser Leu Pro Gln Phe Ser Ala Pro Pro Pro Pro Leu Lys Ile His
                740                 745                 750
Gln Val Gln His Ile Thr Gln Val Ala Pro Pro Thr Pro Pro Pro
                755                 760                 765
Pro Pro Pro Ile Pro Ala Pro Leu Pro Pro Gln Ala Pro Pro Lys
                770                 775                 780
Pro Leu Val Thr Ile Pro Ala Pro Thr Ser Thr Lys Thr Val Ala
                785                 790                 795
Pro Val Val Thr Gln Ala Ala Pro Pro Thr Pro Thr Pro Pro Val
                800                 805                 810
Pro Pro Ala Lys Lys Gln Pro Ala Phe Pro Ala Ser Tyr Ile Pro
                815                 820                 825
Pro Ser Pro Pro Thr Pro Pro Val Pro Val Pro Pro Pro Thr Leu
                830                 835                 840
Pro Lys Gln Gln Ser Phe Cys Ala Lys Pro Pro Pro Ser Pro Leu
                845                 850                 855
Ser Pro Val Pro Ser Val Val Lys Gln Ile Ala Ser Gln Phe Pro
                860                 865                 870
Pro Pro Pro Thr Pro Pro Ala Met Glu Ser Gln Pro Leu Lys Pro
                875                 880                 885
Val Pro Ala Asn Val Ala Pro Gln Ser Pro Pro Ala Val Lys Ala
                890                 895                 900
Lys Pro Lys Trp Gln Pro Ser Ser Ile Pro Val Pro Ser Pro Asp
                905                 910                 915
Phe Pro Pro Pro Pro Pro Glu Ser Ser Leu Val Phe Pro Pro Pro
                920                 925                 930
Pro Pro Ser Pro Val Pro Ala Pro Pro Pro Pro Pro Pro Pro Thr
                935                 940                 945
Ala Ser Pro Thr Pro Asp Lys Ser Gly Ser Pro Gly Lys Lys Thr
                950                 955                 960
Ser Lys Thr Ser Ser Pro Gly Gly Lys Lys Pro Pro Pro Thr Pro
                965                 970                 975
Gln Arg Asn Ser Ser Ile Lys Ser Ser Ser Gly Ala Glu His Pro
                980                 985                 990
Glu Pro Lys Arg Pro Ser Val Asp Ser Leu Val Ser Lys Phe Thr
                995                1000                1005
Pro Pro Ala Glu Ser Gly Ser Pro Ser Lys Glu Thr Leu Pro Pro
               1010                1015                1020
Pro Ala Ala Pro Pro Lys Pro Gly Lys Leu Asn Leu Ser Gly Val
               1025                1030                1035
```

```
Asn Leu Pro Gly Val Leu Gln Gln Gly Cys Val Ser Ala Lys Ala
                1040                1045                1050
Pro Val Leu Ser Gly Arg Gly Lys Asp Ser Val Val Glu Phe Pro
                1055                1060                1065
Ser Pro Pro Ser Asp Ser Asp Phe Pro Pro Pro Pro Pro Glu Thr
                1070                1075                1080
Glu Leu Pro Leu Pro Pro Ile Glu Ile Pro Ala Val Phe Ser Gly
                1085                1090                1095
Asn Thr Ser Pro Lys Val Ala Val Val Asn Pro Gln Pro Gln Gln
                1100                1105                1110
Trp Ser Lys Met Ser Val Lys Lys Ala Pro Pro Pro Thr Arg Pro
                1115                1120                1125
Lys Arg Asn Asp Ser Thr Arg Leu Thr Gln Ala Glu Ile Ser Glu
                1130                1135                1140
Gln Pro Thr Met Ala Thr Val Val Pro Gln Val Pro Thr Ser Pro
                1145                1150                1155
Lys Ser Ser Leu Ser Val Gln Pro Gly Phe Leu Ala Asp Leu Asn
                1160                1165                1170
Arg Thr Leu Gln Arg Lys Ser Ile Thr Arg His Gly Ser Leu Ser
                1175                1180                1185
Ser Arg Met Ser Arg Ala Glu Pro Thr Ala Thr Met Asp Asp Met
                1190                1195                1200
Ala Leu Pro Pro Pro Pro Pro Glu Leu Leu Ser Asp Gln Gln Lys
                1205                1210                1215
Ala Gly Tyr Gly Gly Ser His Ile Ser Gly Tyr Ala Thr Leu Arg
                1220                1225                1230
Arg Gly Pro Pro Pro Ala Pro Pro Lys Arg Asp Gln Asn Thr Lys
                1235                1240                1245
Leu Ser Arg Asp Trp
                1250
```

<210> 10
<211> 621
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 8224777CD1

<400> 10

```
Met Leu Pro Val Asp Gly Glu Glu Arg Lys Ser Glu Gly Ser Asp
  1               5                  10                  15
Thr Val Gly Asp Arg Thr Ser Pro Cys Ala Thr Ser Ser Ala Thr
                20                  25                  30
Leu Lys Asp Leu Glu Val Arg Gly Ser Gly Trp Arg Cys Ser Asp
                35                  40                  45
Pro Ser Gly Gln Pro Ser Asn Leu Leu Leu Gln Val Gly Leu Gly
                50                  55                  60
Ala Pro Leu Pro Ala Glu Thr Ala His Thr His Pro Ser Pro Asn
                65                  70                  75
Asp Arg Ser Leu Tyr Leu Ser Pro His Ser Cys Ser Thr Ser Ser
                80                  85                  90
Ser Leu His Ala Pro Gln Ser Pro Cys Gln Glu Arg Ala Val Val
                95                  100                 105
Leu Asp Ser Thr Ser Val Lys Ile Ser Arg Leu Lys Asn Thr Ile
```

```
                110                 115                 120
Lys Ser Leu Lys Gln Gln Lys Lys Gln Val Glu His Gln Leu Glu
                125                 130                 135
Glu Glu Lys Lys Ala Asn Asn Glu Arg Gln Lys Ala Glu Arg Glu
                140                 145                 150
Leu Glu Val Gln Ile Gln Thr Leu Ile Ile Gln Lys Glu Glu Leu
                155                 160                 165
Asn Thr Asp Leu Tyr His Met Glu Arg Ser Leu Arg Tyr Phe Glu
                170                 175                 180
Glu Glu Ser Lys Asp Leu Ala Val Arg Leu Gln His Ser Leu Gln
                185                 190                 195
Cys Lys Gly Glu Leu Glu Arg Ala Leu Ser Ala Val Ile Ala Thr
                200                 205                 210
Glu Lys Lys Lys Ala Asn Gln Leu Ser Ser Cys Ser Lys Ala His
                215                 220                 225
Thr Glu Trp Glu Leu Glu Gln Ser Leu Gln Asp Gln Ala Leu Leu
                230                 235                 240
Lys Ala Gln Leu Thr Gln Leu Lys Glu Ser Phe Gln Gln Leu Gln
                245                 250                 255
Leu Glu Arg Asp Glu Cys Ala Glu His Ile Glu Gly Glu Arg Ala
                260                 265                 270
Arg Trp His Gln Arg Met Ser Lys Met Ser Gln Glu Ile Cys Thr
                275                 280                 285
Leu Lys Lys Glu Lys Gln Gln Asp Met Arg Arg Val Glu Glu Leu
                290                 295                 300
Glu Arg Ser Leu Ser Lys Leu Lys Asn Gln Met Ala Glu Pro Leu
                305                 310                 315
Pro Pro Glu Pro Pro Ala Val Pro Ser Glu Val Glu Leu Gln His
                320                 325                 330
Val Arg Lys Glu Leu Glu Arg Val Ala Gly Glu Leu Gln Ala Gln
                335                 340                 345
Val Lys Asn Asn Gln His Ile Ser Leu Leu Asn Arg Arg Gln Glu
                350                 355                 360
Glu Arg Ile Arg Glu Gln Glu Glu Arg Leu Arg Lys Gln Glu Glu
                365                 370                 375
Arg Leu Gln Glu Gln His Glu Lys Leu Arg Gln Leu Ala Lys Pro
                380                 385                 390
Gln Ser Val Phe Glu Glu Leu Asn Asn Glu Asn Lys Ser Thr Leu
                395                 400                 405
Gln Leu Glu Gln Gln Val Lys Glu Leu Gln Glu Lys Leu Gly Glu
                410                 415                 420
Val Lys Glu Thr Glu Thr Ser Thr Pro Ser Lys Lys Gly Trp Glu
                425                 430                 435
Ala Gly Ser Ser Leu Leu Gly Gly Glu Val Ser Ser Phe Met Asp
                440                 445                 450
His Leu Lys Glu Lys Ala Asp Leu Ser Glu Leu Val Lys Lys Gln
                455                 460                 465
Glu Leu Arg Phe Ile Gln Tyr Trp Gln Glu Arg Cys His Gln Lys
                470                 475                 480
Ile His His Leu Leu Ser Glu Pro Gly Gly Arg Ala Lys Asp Ala
                485                 490                 495
Ala Leu Gly Gly Gly His His Gln Ala Gly Ala Gln Gly Gly Asp
                500                 505                 510
Glu Gly Glu Ala Ala Gly Ala Ala Ala Asp Gly Ile Ala Ala Tyr
                515                 520                 525
Ser Asn Tyr Asn Asn Gly His Arg Lys Phe Leu Ala Ala Ala His
```

```
                530                 535                 540
Asn Ser Ala Asp Glu Pro Gly Pro Gly Ala Pro Ala Pro Gln Glu
                545                 550                 555
Leu Gly Ala Ala Asp Lys His Gly Asp Leu Arg Glu Val Ser Leu
                560                 565                 570
Thr Ser Ser Ala Gln Gly Glu Ala Arg Glu Asp Pro Leu Leu Asp
                575                 580                 585
Lys Pro Thr Ala Gln Pro Ile Val Gln Asp His Gln Glu His Pro
                590                 595                 600
Gly Leu Gly Ser Asn Cys Cys Val Pro Leu Phe Cys Trp Ala Trp
                605                 610                 615
Leu Pro Arg Arg Arg Arg
                620

<210> 11
<211> 114
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 587394CD1

<400> 11
Met Phe Met Arg Lys Arg Glu Asn Ser Leu Ala Ser Leu Phe Thr
  1               5                  10                  15
Glu Trp Leu Phe Leu Val Ser Ser Arg Cys Ser Arg Thr Asp Val
                20                  25                  30
Arg Phe Gly Ser Ser Glu Ser Thr Lys His Ile Gly Glu Lys Asp
                35                  40                  45
Glu Glu Asp Ser Cys Arg Leu Asp Cys Thr Met Ser Pro Ser Arg
                50                  55                  60
Thr Gln Gln Arg Ala Ala Arg Pro Arg Gly Glu Gly Ser Leu Lys
                65                  70                  75
Gln His Gln Arg Leu His Ser Asn Phe Ser Ser Val Asn Glu Ala
                80                  85                  90
Val Phe Ile Val Ser Phe Ser Leu Asn Thr Asn Leu Lys Thr Thr
                95                  100                 105
Gly Leu Asn Leu Leu Ser His Leu Ala
                110

<210> 12
<211> 527
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1402405CD1

<400> 12
Met Gly Pro Leu Ala Leu Gly Ile Leu Lys Leu Glu His Cys Pro
  1               5                  10                  15
Gln Ala Leu Arg Thr Gln Ala Phe Gln Val Leu Leu Gln Pro Leu
                20                  25                  30
Ala Cys Val Leu Lys Ala Thr Val Gln Ala Pro Gly Pro Pro Gly
```

```
                35                  40                  45
Leu Leu Asp Gly Thr Ala Asp Asp Ala Thr Thr Val Asp Thr Leu
                50                  55                  60
Leu Ala Ser Lys Ser Ser Cys Ala Gly Leu Leu Cys Arg Thr Leu
                65                  70                  75
Ala His Leu Glu Glu Leu Gln Pro Leu Pro Gln Arg Pro Ser Pro
                80                  85                  90
Trp Pro Gln Ala Ser Leu Leu Gly Ala Thr Val Thr Val Leu Arg
                95                  100                 105
Leu Cys Asp Gly Ser Ala Ala Pro Ala Ser Ser Val Gly Gly His
               110                 115                 120
Leu Cys Gly Thr Leu Ala Gly Cys Val Arg Val Gln Arg Ala Ala
               125                 130                 135
Leu Asp Phe Leu Gly Thr Leu Ser Gln Gly Thr Gly Pro Gln Glu
               140                 145                 150
Leu Val Thr Gln Ala Leu Ala Val Leu Leu Glu Cys Leu Glu Ser
               155                 160                 165
Pro Gly Ser Ser Pro Thr Val Leu Lys Lys Ala Phe Gln Ala Thr
               170                 175                 180
Leu Arg Trp Leu Leu Ser Ser Pro Lys Thr Pro Gly Cys Ser Asp
               185                 190                 195
Leu Gly Pro Leu Ile Pro Gln Phe Leu Arg Glu Leu Phe Pro Val
               200                 205                 210
Leu Gln Lys Arg Leu Cys His Pro Cys Trp Glu Val Arg Asp Ser
               215                 220                 225
Ala Leu Glu Phe Leu Thr Gln Leu Ser Arg His Trp Gly Gly Gln
               230                 235                 240
Ala Asp Phe Arg Cys Ala Leu Leu Ala Ser Glu Val Pro Gln Leu
               245                 250                 255
Ala Leu Gln Leu Leu Gln Asp Pro Glu Ser Tyr Val Arg Ala Ser
               260                 265                 270
Ala Val Thr Ala Met Gly Gln Leu Ser Ser Gln Gly Leu His Ala
               275                 280                 285
Pro Thr Ser Pro Glu His Ala Glu Ala Arg Gln Ser Leu Phe Leu
               290                 295                 300
Glu Leu Leu His Ile Leu Ser Val Asp Ser Glu Gly Phe Pro Arg
               305                 310                 315
Arg Ala Val Met Gln Val Phe Thr Glu Trp Leu Arg Asp Gly His
               320                 325                 330
Ala Asp Ala Ala Gln Asp Thr Glu Gln Phe Val Ala Thr Val Leu
               335                 340                 345
Gln Ala Ala Ser Gln Asp Leu Asp Trp Glu Val Arg Ala Gln Gly
               350                 355                 360
Leu Glu Leu Ala Leu Val Phe Leu Gly Gln Thr Leu Gly Pro Pro
               365                 370                 375
Arg Thr His Cys Pro Tyr Ala Val Ala Leu Pro Glu Val Ala Pro
               380                 385                 390
Ala Gln Pro Leu Thr Glu Ala Leu Arg Ala Leu Cys His Val Gly
               395                 400                 405
Leu Phe Asp Phe Ala Phe Cys Ala Leu Phe Asp Cys Asp Arg Pro
               410                 415                 420
Val Ala Gln Lys Ser Cys Asp Leu Leu Leu Phe Leu Arg Asp Lys
               425                 430                 435
Ile Ala Ser Tyr Ser Ser Leu Arg Glu Ala Arg Gly Ser Pro Asn
               440                 445                 450
Thr Ala Ser Ala Glu Ala Thr Leu Pro Arg Trp Arg Ala Gly Glu
```

```
                455                 460                 465
Gln Ala Gln Pro Pro Gly Asp Gln Glu Pro Glu Ala Val Leu Ala
                470                 475                 480
Met Leu Arg Ser Leu Asp Leu Glu Gly Leu Arg Ser Thr Leu Ala
                485                 490                 495
Glu Ser Ser Asp His Val Glu Lys Ser Pro Gln Ser Leu Leu Gln
                500                 505                 510
Asp Met Leu Ala Thr Gly Gly Phe Leu Gln Gly Asp Glu Ala Asp
                515                 520                 525
Cys Tyr
```

<210> 13
<211> 316
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1798468CD1

<400> 13

```
Met Ala Ser Pro Glu His Pro Gly Ser Pro Gly Cys Met Gly Pro
  1               5                  10                  15
Ile Thr Gln Cys Thr Ala Arg Thr Gln Gln Glu Ala Pro Ala Thr
                20                  25                  30
Gly Pro Asp Leu Pro His Pro Gly Pro Asp Gly His Leu Asp Thr
                35                  40                  45
His Ser Gly Leu Ser Ser Asn Ser Ser Met Thr Thr Arg Glu Leu
                50                  55                  60
Gln Gln Tyr Trp Gln Asn Gln Lys Cys Arg Trp Lys His Val Lys
                65                  70                  75
Leu Leu Phe Glu Ile Ala Ser Ala Arg Ile Glu Glu Arg Lys Val
                80                  85                  90
Ser Lys Phe Val Val Tyr Gln Ile Ile Val Ile Gln Thr Gly Ser
                95                 100                 105
Phe Asp Asn Asn Lys Ala Val Leu Glu Arg Arg Tyr Ser Asp Phe
               110                 115                 120
Ala Lys Leu Gln Lys Ala Leu Leu Lys Thr Phe Arg Glu Glu Ile
               125                 130                 135
Glu Asp Val Glu Phe Pro Arg Lys His Leu Thr Gly Asn Phe Ala
               140                 145                 150
Glu Glu Met Ile Cys Glu Arg Arg Arg Ala Leu Gln Glu Tyr Leu
               155                 160                 165
Gly Leu Leu Tyr Ala Ile Arg Cys Val Arg Arg Ser Arg Glu Phe
               170                 175                 180
Leu Asp Phe Leu Thr Arg Pro Glu Leu Arg Glu Ala Phe Gly Cys
               185                 190                 195
Leu Arg Ala Gly Gln Tyr Pro Arg Ala Leu Glu Leu Leu Leu Arg
               200                 205                 210
Val Leu Pro Leu Gln Glu Lys Leu Thr Ala His Cys Pro Ala Ala
               215                 220                 225
Ala Val Pro Ala Leu Cys Ala Val Leu Leu Cys His Arg Asp Leu
               230                 235                 240
Asp Arg Pro Ala Glu Ala Phe Ala Ala Gly Glu Arg Ala Leu Gln
               245                 250                 255
```

```
Arg Leu Gln Ala Arg Glu Gly His Arg Tyr Tyr Ala Pro Leu Leu
                260                 265                 270
Asp Ala Met Val Arg Leu Ala Tyr Ala Leu Gly Lys Asp Phe Val
                275                 280                 285
Thr Leu Gln Glu Arg Leu Glu Glu Ser Gln Leu Arg Arg Pro Thr
                290                 295                 300
Pro Arg Gly Ile Thr Leu Lys Glu Leu Thr Val Arg Glu Tyr Leu
                305                 310                 315
His
```

<210> 14
<211> 659
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3189084CD1

<400> 14

```
Met Gly Asp Cys Ala Glu Ile Lys Ser Gln Phe Arg Thr Arg Glu
  1               5                  10                  15
Gly Phe Tyr Lys Leu Leu Pro Gly Asp Gly Ala Ala Arg Arg Ser
                20                  25                  30
Gly Pro Ala Ser Ala Gln Thr Pro Val Pro Pro Gln Pro Pro Gln
                35                  40                  45
Pro Pro Pro Gly Pro Ala Ser Ala Ser Gly Pro Gly Ala Ala Gly
                50                  55                  60
Pro Ala Ser Ser Pro Pro Pro Ala Gly Pro Gly Pro Gly Pro Ala
                65                  70                  75
Leu Pro Ala Val Arg Leu Ser Leu Val Arg Leu Gly Glu Pro Asp
                80                  85                  90
Ser Ala Gly Ala Gly Glu Pro Pro Ala Thr Pro Ala Gly Leu Gly
                95                 100                 105
Ser Gly Gly Asp Arg Val Cys Phe Asn Leu Gly Arg Glu Leu Tyr
               110                 115                 120
Phe Tyr Pro Gly Cys Cys Arg Arg Gly Ser Gln Arg Ser Ile Asp
               125                 130                 135
Leu Asn Lys Pro Ile Asp Lys Arg Ile Tyr Lys Gly Thr Gln Pro
               140                 145                 150
Thr Cys His Asp Phe Asn Gln Phe Thr Ala Ala Thr Glu Thr Ile
               155                 160                 165
Ser Leu Leu Val Gly Phe Ser Ala Gly Gln Val Gln Tyr Leu Asp
               170                 175                 180
Leu Ile Lys Lys Asp Thr Ser Lys Leu Phe Asn Glu Glu Arg Leu
               185                 190                 195
Ile Asp Lys Thr Lys Val Thr Tyr Leu Lys Trp Leu Pro Glu Ser
               200                 205                 210
Glu Ser Leu Phe Leu Ala Ser His Ala Ser Gly His Leu Tyr Leu
               215                 220                 225
Tyr Asn Val Ser His Pro Cys Ala Ser Ala Pro Pro Gln Tyr Ser
               230                 235                 240
Leu Leu Lys Gln Gly Glu Gly Phe Ser Val Tyr Ala Ala Lys Ser
               245                 250                 255
Lys Ala Pro Arg Asn Pro Leu Ala Lys Trp Ala Val Gly Glu Gly
```

```
                260                 265                 270
Pro Leu Asn Glu Phe Ala Phe Ser Pro Asp Gly Arg His Leu Ala
                275                 280                 285
Cys Val Ser Gln Asp Gly Cys Leu Arg Val Phe His Phe Asp Ser
                290                 295                 300
Met Leu Leu Arg Gly Leu Met Lys Ser Tyr Phe Gly Gly Leu Leu
                305                 310                 315
Cys Val Cys Trp Ser Pro Asp Gly Arg Tyr Val Val Thr Gly Gly
                320                 325                 330
Glu Asp Asp Leu Val Thr Val Trp Ser Phe Thr Glu Gly Arg Val
                335                 340                 345
Val Ala Arg Gly His Gly His Lys Ser Trp Val Asn Ala Val Ala
                350                 355                 360
Phe Asp Pro Tyr Thr Thr Arg Ala Glu Glu Ala Ala Thr Ala Ala
                365                 370                 375
Gly Ala Asp Gly Glu Arg Ser Gly Glu Glu Glu Glu Glu Glu Pro
                380                 385                 390
Glu Ala Ala Gly Thr Gly Ser Ala Gly Gly Ala Pro Leu Ser Pro
                395                 400                 405
Leu Pro Lys Ala Gly Ser Ile Thr Tyr Arg Phe Gly Ser Ala Gly
                410                 415                 420
Gln Asp Thr Gln Phe Cys Leu Trp Asp Leu Thr Glu Asp Val Leu
                425                 430                 435
Tyr Pro His Pro Pro Leu Ala Arg Thr Arg Thr Leu Pro Gly Thr
                440                 445                 450
Pro Gly Thr Thr Pro Pro Ala Ala Ser Ser Ser Arg Gly Gly Glu
                455                 460                 465
Pro Gly Pro Gly Pro Leu Pro Arg Ser Leu Ser Arg Ser Asn Ser
                470                 475                 480
Leu Pro His Pro Ala Gly Gly Gly Lys Ala Gly Gly Pro Gly Val
                485                 490                 495
Ala Ala Glu Pro Gly Thr Pro Phe Ser Ile Gly Arg Phe Ala Thr
                500                 505                 510
Leu Thr Leu Gln Glu Arg Arg Asp Arg Gly Ala Glu Lys Glu His
                515                 520                 525
Lys Arg Tyr His Ser Leu Gly Asn Ile Ser Arg Gly Gly Ser Gly
                530                 535                 540
Gly Ser Gly Ser Gly Gly Glu Lys Pro Ser Gly Pro Val Pro Arg
                545                 550                 555
Ser Arg Leu Asp Pro Ala Lys Val Leu Gly Thr Ala Leu Cys Pro
                560                 565                 570
Arg Ile His Glu Val Pro Leu Leu Glu Pro Leu Val Cys Lys Lys
                575                 580                 585
Ile Ala Gln Glu Arg Leu Thr Val Leu Leu Phe Leu Glu Asp Cys
                590                 595                 600
Ile Ile Thr Ala Cys Gln Glu Gly Leu Ile Cys Thr Trp Ala Arg
                605                 610                 615
Pro Gly Lys Ala Phe Thr Asp Glu Glu Thr Glu Ala Gln Thr Gly
                620                 625                 630
Glu Gly Ser Trp Pro Arg Ser Pro Ser Lys Ser Val Val Glu Gly
                635                 640                 645
Ile Ser Ser Gln Pro Gly Asn Ser Pro Ser Gly Thr Val Val
                650                 655
```

<210> 15
<211> 446

<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5580384CD1

<400> 15

```
Met Asn Val Thr Pro Glu Val Lys Ser Arg Gly Met Lys Phe Ala
  1               5                  10                  15
Glu Glu Gln Leu Leu Lys His Gly Trp Thr Gln Gly Lys Gly Leu
                 20                  25                  30
Gly Arg Lys Glu Asn Gly Ile Thr Gln Ala Leu Arg Val Thr Leu
                 35                  40                  45
Lys Gln Asp Thr His Gly Val Gly His Asp Pro Ala Lys Glu Phe
                 50                  55                  60
Thr Asn His Trp Trp Asn Glu Leu Phe Asn Lys Thr Ala Ala Asn
                 65                  70                  75
Leu Val Val Glu Thr Gly Gln Asp Gly Val Gln Ile Arg Ser Leu
                 80                  85                  90
Ser Lys Glu Thr Thr Arg Tyr Asn His Pro Lys Pro Asn Leu Leu
                 95                 100                 105
Tyr Gln Lys Phe Val Lys Met Ala Thr Leu Thr Ser Gly Gly Glu
                110                 115                 120
Lys Pro Asn Lys Asp Leu Glu Ser Cys Ser Asp Asp Asp Asn Gln
                125                 130                 135
Gly Ser Lys Ser Pro Lys Ile Leu Thr Asp Glu Met Leu Leu Gln
                140                 145                 150
Ala Cys Glu Gly Arg Thr Ala His Lys Ala Ala Arg Leu Gly Ile
                155                 160                 165
Thr Met Lys Ala Lys Leu Ala Arg Leu Glu Ala Gln Glu Gln Ala
                170                 175                 180
Phe Leu Ala Arg Leu Lys Gly Gln Asp Pro Gly Ala Pro Gln Leu
                185                 190                 195
Gln Ser Glu Ser Lys Pro Pro Lys Lys Lys Lys Lys Arg Arg
                200                 205                 210
Gln Lys Glu Glu Glu Glu Ala Thr Ala Ser Glu Arg Asn Asp Ala
                215                 220                 225
Asp Glu Lys His Pro Glu His Ala Glu Gln Asn Ile Arg Lys Ser
                230                 235                 240
Lys Lys Lys Lys Arg Arg His Gln Glu Gly Lys Val Ser Asp Glu
                245                 250                 255
Arg Glu Gly Thr Thr Lys Gly Asn Glu Lys Glu Asp Ala Ala Gly
                260                 265                 270
Thr Ser Gly Leu Gly Glu Leu Asn Ser Arg Glu Gln Thr Asn Gln
                275                 280                 285
Ser Leu Arg Lys Gly Lys Lys Lys Lys Arg Trp His His Glu Glu
                290                 295                 300
Glu Lys Met Gly Val Leu Glu Glu Gly Gly Lys Gly Lys Glu Ala
                305                 310                 315
Ala Gly Ser Val Arg Thr Glu Glu Val Glu Ser Arg Ala Tyr Ala
                320                 325                 330
Asp Pro Cys Ser Arg Arg Lys Lys Arg Gln Gln Gln Glu Glu Glu
                335                 340                 345
Asp Leu Asn Leu Glu Asp Arg Gly Glu Glu Thr Val Leu Gly Gly
                350                 355                 360
```

```
Gly Thr Arg Glu Ala Glu Ser Arg Ala Cys Ser Asp Gly Arg Ser
                365                 370                 375
Arg Lys Ser Lys Lys Lys Arg Gln Gln His Gln Glu Glu Glu Asp
                380                 385                 390
Ile Leu Asp Val Arg Asp Glu Lys Asp Ser Gly Ala Arg Glu Ala
                395                 400                 405
Glu Ser Arg Ala His Thr Gly Ser Ser Ser Arg Gly Lys Arg Lys
                410                 415                 420
Arg Gln Gln His Pro Lys Lys Glu Arg Ala Gly Val Ser Thr Val
                425                 430                 435
Gln Lys Ala Lys Lys Lys Gln Lys Lys Arg Asp
                440                 445
```

<210> 16
<211> 364
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5158619CD1

<400> 16

```
Met Thr Glu Lys Glu Val Leu Glu Ser Pro Lys Pro Ser Phe Pro
  1               5                  10                  15
Ala Glu Thr Arg Gln Ser Gly Leu Gln Arg Leu Lys Gln Leu Leu
                20                  25                  30
Arg Lys Gly Ser Thr Gly Thr Lys Glu Met Glu Leu Pro Pro Glu
                35                  40                  45
Pro Gln Ala Asn Gly Glu Ala Val Gly Ala Gly Gly Gly Pro Ile
                50                  55                  60
Tyr Tyr Ile Tyr Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu
                65                  70                  75
Glu Pro Pro Pro Glu Pro Pro Lys Leu Val Asn Asp Lys Pro His
                80                  85                  90
Lys Phe Lys Asp His Phe Phe Lys Lys Pro Lys Phe Cys Asp Val
                95                 100                 105
Cys Ala Arg Met Ile Val Leu Asn Asn Lys Phe Gly Leu Arg Cys
               110                 115                 120
Lys Asn Cys Lys Thr Asn Ile His Glu His Cys Gln Ser Tyr Val
               125                 130                 135
Glu Met Gln Arg Cys Phe Gly Lys Ile Pro Pro Gly Phe His Arg
               140                 145                 150
Ala Tyr Ser Ser Pro Leu Tyr Ser Asn Gln Gln Tyr Ala Cys Val
               155                 160                 165
Lys Asp Leu Ser Ala Ala Asn Arg Asn Asp Pro Val Phe Glu Thr
               170                 175                 180
Leu Arg Thr Gly Val Ile Met Ala Asn Lys Glu Arg Lys Lys Gly
               185                 190                 195
Gln Ala Asp Lys Lys Asn Pro Val Ala Ala Met Met Glu Glu Glu
               200                 205                 210
Pro Glu Ser Ala Arg Pro Glu Glu Gly Lys Pro Gln Asp Gly Asn
               215                 220                 225
Pro Glu Gly Asp Lys Lys Ala Glu Lys Lys Thr Pro Asp Asp Lys
               230                 235                 240
His Lys Gln Pro Gly Phe Gln Gln Ser His Tyr Phe Val Ala Leu
```

```
                245                 250                 255
Tyr Arg Phe Lys Ala Leu Glu Lys Asp Asp Leu Asp Phe Pro Pro
                260                 265                 270
Gly Glu Lys Ile Thr Val Ile Asp Asp Ser Asn Glu Glu Trp Trp
                275                 280                 285
Arg Gly Lys Ile Gly Glu Lys Val Gly Phe Phe Pro Pro Asn Phe
                290                 295                 300
Ile Ile Arg Val Arg Ala Gly Glu Arg Val His Arg Val Thr Arg
                305                 310                 315
Ser Phe Val Gly Asn Arg Glu Ile Gly Gln Ile Thr Leu Lys Lys
                320                 325                 330
Asp Gln Ile Val Val Gln Lys Gly Asp Glu Ala Gly Gly Tyr Val
                335                 340                 345
Lys Val Tyr Thr Gly Arg Lys Val Gly Leu Phe Pro Thr Asp Phe
                350                 355                 360
Leu Glu Glu Ile
```

<210> 17
<211> 91
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2792745CD1

<400> 17

```
Met Glu Phe Cys Ser Phe Ala Gln Thr Gly Val Gln Arg Arg Asp
  1               5                  10                  15
Arg Gly Ser Leu Gln Pro Pro Pro Pro Lys Phe Lys Gln Phe Ser
                 20                  25                  30
His Leu Ser Leu Leu Ser Ser Trp Asp Tyr Tyr Arg His Pro Pro
                 35                  40                  45
Ser Arg Pro Asp Asn Phe Trp Ile Phe Val Val Met Gly Leu His
                 50                  55                  60
His Val Gly Leu Ala Gly Leu Gln Leu Leu Thr Ser Ser Asp Pro
                 65                  70                  75
Pro Thr Ser Thr Ser Gln Ser Ala Gly Ile Thr Ser Val Asn His
                 80                  85                  90
Arg
```

<210> 18
<211> 116
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2827678CD1

<400> 18

```
Met Pro Pro Ser Ser Ala Asn Phe Phe Cys Leu Phe Val Phe Val
  1               5                  10                  15
Phe Val Leu Arg Trp Ser Phe Val Leu Val Ala Gln Ala Gly Val
```

```
                20                  25                  30
Gln Trp Cys Arg Leu Gly Ser Pro Gln Pro Leu Pro Pro Arg Phe
                35                  40                  45
Lys Arg Phe Ser Cys Leu Thr Leu Pro Ser Ser Trp Asp Tyr Arg
                50                  55                  60
Cys Leu Pro Pro Arg Pro Ala Asn Phe Phe Val Phe Leu Val Glu
                65                  70                  75
Thr Gly Phe His His Ile Ala Gln Ala Gly Phe Gln Leu Leu Thr
                80                  85                  90
Ser Gly Asp Pro Pro Ala Leu Ser Ser Gln Ser Ala Gly Ile Thr
                95                  100                 105
Gly Val Ser His Cys Ala Trp Pro Thr Phe Leu
                110                 115

<210> 19
<211> 684
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 790257CD1

<400> 19
Met Ser Ala Ser Gly Val Leu Ser Phe Thr Gln Gln Gly Trp Glu
  1                 5                  10                  15
Gln Val Leu Ala Lys Val Lys Arg Ala Val Val Tyr Leu Asp Ala
                20                  25                  30
Ala Cys Ala Glu Ser Leu His Trp Gly Cys Gly Ser Thr Arg Leu
                35                  40                  45
Leu Glu Ala Val Gly Gly Pro Asp Cys His Leu Arg Glu Phe Glu
                50                  55                  60
Pro Asp Ala Ile Gly Gly Gly Ala Lys Gln Pro Lys Ala Val Phe
                65                  70                  75
Val Leu Ser Cys Leu Leu Lys Gly Arg Thr Val Glu Ile Leu Arg
                80                  85                  90
Asp Ile Ile Cys Arg Ser His Phe Gln Tyr Cys Val Val Val Thr
                95                  100                 105
Thr Val Ser His Ala Val His Leu Thr Ala Asn His Val Pro Ala
                110                 115                 120
Ala Ala Ala Ala Glu Met Glu Gly Gln Gln Pro Val Phe Glu Gln
                125                 130                 135
Leu Glu Glu Lys Leu Cys Glu Trp Met Gly Asn Met Asn Tyr Thr
                140                 145                 150
Ala Glu Val Phe His Val Pro Leu Leu Leu Ala Pro Val Ala Pro
                155                 160                 165
His Phe Ala Leu Thr Pro Ala Phe Ala Ser Leu Phe Pro Leu Leu
                170                 175                 180
Pro Gln Asp Val His Leu Leu Asn Ser Ala Arg Pro Asp Lys Arg
                185                 190                 195
Lys Leu Gly Ser Leu Gly Asp Val Asp Ser Thr Thr Leu Thr Pro
                200                 205                 210
Glu Leu Leu Leu Gln Ile Arg Cys Leu Val Ser Gly Leu Ser Ser
                215                 220                 225
Leu Cys Glu His Leu Gly Val Arg Glu Glu Cys Phe Ala Val Gly
                230                 235                 240
```

```
Ser Leu Ser Gln Val Ile Ala Ala Asp Leu Ala Asn Tyr Ala Pro
                245                 250                 255
Ala Lys Asn Arg Lys Lys Thr Ala Ala Gly Arg Ala Ser Val Val
                260                 265                 270
Phe Val Asp Arg Thr Leu Asp Leu Thr Gly Ala Val Gly His His
                275                 280                 285
Gly Asp Asn Leu Val Glu Lys Ile Ile Ser Ala Leu Pro Gln Leu
                290                 295                 300
Pro Gly His Thr Asn Asp Val Met Val Asn Met Ile Ala Leu Thr
                305                 310                 315
Ala Leu His Thr Glu Glu Glu Asn Tyr Asn Val Val Ala Pro Gly
                320                 325                 330
Cys Leu Ser Gln Ser Ser Asp Thr Thr Ala Lys Ala Leu Trp Glu
                335                 340                 345
Ala Leu Leu Asn Thr Lys His Lys Glu Ala Val Met Glu Val Arg
                350                 355                 360
Arg His Leu Val Glu Ala Ala Ser Arg Glu Asn Leu Pro Ile Lys
                365                 370                 375
Met Ser Met Gly Arg Val Thr Pro Gly Gln Leu Met Ser Tyr Ile
                380                 385                 390
Gln Leu Phe Lys Asn Asn Leu Lys Ala Leu Met Asn His Cys Gly
                395                 400                 405
Leu Leu Gln Leu Gly Leu Ala Thr Ala Gln Thr Leu Lys His Pro
                410                 415                 420
Gln Thr Ala Lys Trp Asp Asn Phe Leu Ala Phe Glu Arg Leu Leu
                425                 430                 435
Leu Gln Ser Ile Gly Glu Ser Ala Met Ser Val Val Leu Asn Gln
                440                 445                 450
Leu Leu Pro Met Ile Lys Pro Val Thr Gln Arg Thr Asn Glu Asp
                455                 460                 465
Tyr Ser Pro Glu Glu Leu Leu Ile Leu Leu Ile Tyr Ile Tyr Ser
                470                 475                 480
Val Thr Gly Glu Leu Thr Val Asp Lys Asp Leu Cys Glu Ala Glu
                485                 490                 495
Glu Lys Val Lys Lys Ala Leu Ala Gln Val Phe Cys Glu Glu Ser
                500                 505                 510
Gly Ser Ser Pro Leu Leu Gln Lys Ile Thr Asp Trp Asp Ser Ser
                515                 520                 525
Ile Asn Leu Thr Phe His Lys Ser Lys Ile Ala Val Asp Glu Leu
                530                 535                 540
Phe Thr Ser Leu Arg Asp Ile Ala Gly Ala Arg Ser Leu Leu Lys
                545                 550                 555
Gln Phe Lys Ser Val Tyr Val Pro Gly Asn His Thr His Gln Ala
                560                 565                 570
Ser Tyr Lys Pro Leu Leu Lys Gln Val Val Glu Glu Ile Phe His
                575                 580                 585
Pro Glu Arg Pro Asp Ser Val Asp Ile Glu His Met Ser Ser Gly
                590                 595                 600
Leu Thr Asp Leu Leu Lys Thr Gly Phe Ser Met Phe Met Lys Val
                605                 610                 615
Ser Arg Pro His Pro Ser Asp Tyr Pro Leu Leu Ile Leu Phe Val
                620                 625                 630
Val Gly Gly Val Thr Val Ser Glu Val Lys Met Val Lys Asp Leu
                635                 640                 645
Val Ala Ser Leu Lys Pro Gly Thr Gln Val Ile Val Leu Ser Thr
                650                 655                 660
```

```
Arg Leu Leu Lys Pro Leu Asn Ile Pro Glu Leu Leu Phe Ala Thr
                665                 670                 675
Asp Arg Leu His Pro Asp Leu Gly Phe
                680

<210> 20
<211> 344
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2617345CD1

<400> 20
Met Asn Leu Leu Pro Cys Asn Pro His Gly Asn Gly Leu Leu Tyr
  1               5                  10                  15
Ala Gly Phe Asn Gln Asp His Gly Cys Phe Ala Cys Gly Met Glu
                20                  25                  30
Asn Gly Phe Arg Val Tyr Asn Thr Asp Pro Leu Lys Glu Lys Glu
                35                  40                  45
Lys Gln Glu Phe Leu Glu Gly Gly Val Gly His Val Glu Met Leu
                50                  55                  60
Phe Arg Cys Asn Tyr Leu Ala Leu Val Gly Gly Gly Lys Lys Pro
                65                  70                  75
Lys Tyr Pro Pro Asn Lys Val Met Ile Trp Asp Asp Leu Lys Lys
                80                  85                  90
Lys Thr Val Ile Glu Ile Glu Phe Ser Thr Glu Val Lys Ala Val
                95                 100                 105
Lys Leu Arg Arg Asp Arg Ile Val Val Val Leu Asp Ser Met Ile
               110                 115                 120
Lys Val Phe Thr Phe Thr His Asn Pro His Gln Leu His Val Phe
               125                 130                 135
Glu Thr Cys Tyr Asn Pro Lys Gly Leu Cys Val Leu Cys Pro Asn
               140                 145                 150
Ser Asn Asn Ser Leu Leu Ala Phe Pro Gly Thr His Thr Gly His
               155                 160                 165
Val Gln Leu Val Asp Leu Ala Ser Thr Glu Lys Pro Pro Val Asp
               170                 175                 180
Ile Pro Ala His Glu Gly Val Leu Ser Cys Ile Ala Leu Asn Leu
               185                 190                 195
Gln Gly Thr Arg Ile Ala Thr Ala Ser Glu Lys Gly Thr Leu Ile
               200                 205                 210
Arg Ile Phe Asp Thr Ser Ser Gly His Leu Ile Gln Glu Leu Arg
               215                 220                 225
Arg Gly Ser Gln Ala Ala Asn Ile Tyr Cys Ile Asn Phe Asn Gln
               230                 235                 240
Asp Ala Ser Leu Ile Cys Val Ser Ser Asp His Gly Thr Val His
               245                 250                 255
Ile Phe Ala Ala Glu Asp Pro Lys Arg Asn Lys Gln Ser Ser Leu
               260                 265                 270
Ala Ser Ala Ser Phe Leu Pro Lys Tyr Phe Ser Ser Lys Trp Ser
               275                 280                 285
Phe Ser Lys Phe Gln Val Pro Ser Gly Ser Pro Cys Ile Cys Ala
               290                 295                 300
Phe Gly Thr Glu Pro Asn Ala Val Ile Ala Ile Cys Ala Asp Gly
```

```
                305                 310                 315
Ser Tyr Tyr Lys Phe Leu Phe Asn Pro Lys Gly Glu Cys Ile Arg
                320                 325                 330
Asp Val Tyr Ala Gln Phe Leu Glu Met Thr Asp Asp Lys Leu
                335                 340

<210> 21
<211> 95
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3254666CD1

<400> 21
Met Gly Cys Met Lys Ser Lys Gln Thr Phe Pro Phe Pro Thr Ile
  1               5                  10                  15
Tyr Glu Gly Glu Lys Gln His Glu Ser Glu Glu Pro Phe Met Pro
                20                  25                  30
Glu Glu Arg Cys Leu Pro Arg Met Ala Ser Pro Val Asn Val Lys
                35                  40                  45
Glu Glu Val Lys Glu Pro Pro Gly Thr Asn Thr Val Ile Leu Glu
                50                  55                  60
Tyr Ala His Arg Leu Ser Gln Asp Ile Leu Cys Asp Ala Leu Gln
                65                  70                  75
Gln Trp Ala Cys Asn Asn Ile Lys Tyr His Asp Ile Pro Tyr Ile
                80                  85                  90
Glu Ser Glu Gly Pro
                95

<210> 22
<211> 410
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4159378CD1

<400> 22
Met Pro Tyr Ser Thr Asn Lys Glu Leu Ile Leu Gly Ile Met Val
  1               5                  10                  15
Gly Thr Ala Gly Ile Ser Leu Leu Leu Leu Trp Tyr His Lys Val
                20                  25                  30
Arg Lys Pro Gly Ile Ala Met Lys Leu Pro Glu Phe Leu Ser Leu
                35                  40                  45
Gly Asn Thr Phe Asn Ser Ile Thr Leu Gln Asp Glu Ile His Asp
                50                  55                  60
Asp Gln Gly Thr Thr Val Ile Phe Gln Glu Arg Gln Leu Gln Ile
                65                  70                  75
Leu Glu Lys Leu Asn Glu Leu Leu Thr Asn Met Glu Glu Leu Lys
                80                  85                  90
Glu Glu Ile Arg Phe Leu Lys Glu Ala Ile Pro Lys Leu Glu Glu
                95                 100                 105
Tyr Ile Gln Asp Glu Leu Gly Gly Lys Ile Thr Val His Lys Ile
```

```
                110                 115                 120
Ser Pro Gln His Arg Ala Arg Lys Arg Arg Leu Pro Thr Ile Gln
                125                 130                 135
Ser Ser Ala Thr Ser Asn Ser Ser Glu Glu Ala Glu Ser Glu Gly
                140                 145                 150
Gly Tyr Ile Thr Ala Asn Thr Asp Thr Glu Glu Gln Ser Phe Pro
                155                 160                 165
Val Pro Lys Ala Phe Asn Thr Arg Val Glu Glu Leu Asn Leu Asp
                170                 175                 180
Val Leu Leu Gln Lys Val Asp His Leu Arg Met Ser Glu Ser Gly
                185                 190                 195
Lys Ser Glu Ser Phe Glu Leu Leu Arg Asp His Lys Glu Lys Phe
                200                 205                 210
Arg Asp Glu Ile Glu Phe Met Trp Arg Phe Ala Arg Ala Tyr Gly
                215                 220                 225
Asp Met Tyr Glu Leu Ser Thr Asn Thr Gln Glu Lys Lys His Tyr
                230                 235                 240
Ala Asn Ile Gly Lys Thr Leu Ser Glu Arg Ala Ile Asn Arg Ala
                245                 250                 255
Pro Met Asn Gly His Cys His Leu Trp Tyr Ala Val Leu Cys Gly
                260                 265                 270
Tyr Val Ser Glu Phe Glu Gly Leu Gln Asn Lys Ile Asn Tyr Gly
                275                 280                 285
His Leu Phe Lys Glu His Leu Asp Ile Ala Ile Lys Leu Leu Pro
                290                 295                 300
Glu Glu Pro Phe Leu Tyr Tyr Leu Lys Gly Arg Tyr Cys Tyr Thr
                305                 310                 315
Val Ser Lys Leu Ser Trp Ile Glu Lys Lys Met Ala Ala Thr Leu
                320                 325                 330
Phe Gly Lys Ile Pro Ser Ser Thr Val Gln Glu Ala Leu His Asn
                335                 340                 345
Phe Leu Lys Ala Glu Glu Leu Cys Pro Gly Tyr Ser Asn Pro Asn
                350                 355                 360
Tyr Met Tyr Leu Ala Lys Cys Tyr Thr Asp Leu Glu Glu Asn Gln
                365                 370                 375
Asn Ala Leu Lys Phe Cys Asn Leu Ala Leu Leu Leu Pro Thr Val
                380                 385                 390
Thr Lys Glu Asp Lys Glu Ala Gln Lys Glu Met Gln Lys Ile Met
                395                 400                 405
Thr Ser Leu Lys Arg
                410

<210> 23
<211> 616
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4317538CD1

<400> 23
Met Lys Cys Ala Leu Phe Leu Arg Val Lys Ile Leu Gln Arg Val
  1               5                  10                  15
Cys Arg Ser Arg Tyr Ile Val His Ala Thr Cys Asp Ser Thr Ala
                 20                  25                  30
```

```
Ala Met Ser Gly Ile Leu Lys Arg Lys Phe Glu Glu Val Asp Gly
                35                  40                  45
Ser Ser Pro Cys Ser Ser Val Arg Glu Ser Asp Asp Glu Val Ser
                50                  55                  60
Ser Ser Glu Ser Ala Asp Ser Gly Asp Ser Val Asn Pro Ser Thr
                65                  70                  75
Ser Ser His Phe Thr Pro Ser Ser Ile Leu Lys Arg Glu Lys Arg
                80                  85                  90
Leu Arg Thr Lys Asn Val His Phe Ser Cys Val Thr Val Tyr Tyr
                95                 100                 105
Phe Thr Arg Arg Gln Gly Phe Thr Ser Val Pro Ser Gln Gly Gly
               110                 115                 120
Ser Thr Leu Gly Met Ser Ser Arg His Asn Ser Val Arg Gln Tyr
               125                 130                 135
Thr Leu Gly Glu Phe Ala Arg Glu Gln Glu Arg Leu His Arg Glu
               140                 145                 150
Met Leu Arg Glu His Leu Arg Glu Glu Lys Leu Asn Ser Leu Lys
               155                 160                 165
Leu Lys Met Thr Lys Asn Gly Thr Val Glu Ser Glu Glu Ala Ser
               170                 175                 180
Thr Leu Thr Leu Asp Asp Ile Ser Asp Asp Asp Ile Asp Leu Asp
               185                 190                 195
Asn Thr Glu Val Asp Glu Tyr Phe Phe Leu Gln Pro Leu Pro Thr
               200                 205                 210
Lys Lys Arg Arg Ala Leu Leu Arg Ala Ser Gly Val Lys Lys Ile
               215                 220                 225
Asp Val Glu Glu Lys His Glu Leu Arg Ala Ile Arg Leu Ser Arg
               230                 235                 240
Glu Asp Cys Gly Cys Asp Cys Arg Val Phe Cys Asp Pro Asp Thr
               245                 250                 255
Cys Thr Cys Ser Leu Ala Gly Ile Lys Cys Gln Val Asp Arg Met
               260                 265                 270
Ser Phe Pro Cys Gly Cys Thr Lys Glu Gly Cys Ser Asn Thr Ala
               275                 280                 285
Gly Arg Ile Glu Phe Asn Pro Ile Arg Val Arg Thr His Phe Leu
               290                 295                 300
His Thr Ile Met Lys Leu Glu Leu Glu Lys Asn Arg Glu Gln Gln
               305                 310                 315
Ile Pro Thr Leu Asn Gly Cys His Ser Glu Ile Ser Ala His Ser
               320                 325                 330
Ser Ser Met Gly Pro Val Ala His Ser Val Glu Tyr Ser Ile Ala
               335                 340                 345
Asp Ser Phe Glu Ile Glu Thr Glu Pro Gln Ala Ala Val Leu His
               350                 355                 360
Leu Gln Ser Ala Glu Glu Leu Asp Cys Gln Gly Glu Glu Glu Glu
               365                 370                 375
Glu Glu Glu Asp Gly Ser Ser Phe Cys Gly Gly Val Thr Asp Ser
               380                 385                 390
Ser Thr Gln Ser Leu Ala Pro Ser Glu Ser Asp Glu Glu Glu Glu
               395                 400                 405
Glu Glu Glu Glu Glu Glu Glu Glu Glu Asp Asp Asp Asp Asp Lys
               410                 415                 420
Gly Asp Gly Phe Val Glu Gly Leu Gly Thr His Ala Glu Val Val
               425                 430                 435
Pro Leu Pro Ser Val Leu Cys Tyr Ser Asp Gly Thr Ala Val His
               440                 445                 450
```

```
Glu Ser His Ala Lys Asn Ala Ser Phe Tyr Ala Asn Ser Ser Thr
                455                 460                 465
Leu Tyr Tyr Gln Ile Asp Ser His Ile Pro Gly Thr Pro Asn Gln
                470                 475                 480
Ile Ser Glu Asn Tyr Ser Glu Arg Asp Thr Val Lys Asn Gly Thr
                485                 490                 495
Leu Ser Leu Val Pro Tyr Thr Met Thr Pro Glu Gln Phe Val Asp
                500                 505                 510
Tyr Ala Arg Gln Ala Glu Glu Ala Tyr Gly Ala Ser His Tyr Pro
                515                 520                 525
Ala Ala Asn Pro Ser Val Ile Val Cys Cys Ser Ser Ser Glu Asn
                530                 535                 540
Asp Ser Gly Val Pro Cys Asn Ser Leu Tyr Pro Glu His Arg Ser
                545                 550                 555
Asn His Pro Gln Val Glu Ser His Ser Tyr Leu Lys Gly Pro Ser
                560                 565                 570
Gln Glu Gly Phe Val Ser Ala Leu Asn Gly Asp Ser His Ile Ser
                575                 580                 585
Glu His Pro Ala Glu Asn Ser Leu Ser Leu Ala Glu Lys Ser Ile
                590                 595                 600
Leu His Glu Glu Cys Ile Lys Ser Pro Val Val Glu Thr Val Pro
                605                 610                 615
Val
```

<210> 24
<211> 392
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1881010CD1

<400> 24

```
Met Ala Thr Ala Ala Gln Gly Pro Leu Ser Leu Leu Trp Gly Trp
  1               5                  10                  15
Leu Trp Ser Glu Arg Phe Trp Leu Pro Glu Asn Val Ser Trp Ala
                20                  25                  30
Asp Leu Glu Gly Pro Ala Asp Gly Tyr Gly Tyr Pro Arg Gly Arg
                35                  40                  45
His Ile Leu Ser Val Phe Pro Leu Ala Ala Gly Ile Phe Phe Val
                50                  55                  60
Arg Leu Leu Phe Glu Arg Phe Ile Ala Lys Pro Cys Ala Leu Cys
                65                  70                  75
Ile Gly Ile Glu Asp Ser Gly Pro Tyr Gln Ala Gln Pro Asn Ala
                80                  85                  90
Ile Leu Glu Lys Val Phe Ile Ser Ile Thr Lys Tyr Pro Asp Lys
                95                 100                 105
Lys Arg Leu Glu Gly Leu Ser Lys Gln Leu Asp Trp Asn Val Arg
               110                 115                 120
Lys Ile Gln Cys Trp Phe Arg His Arg Arg Asn Gln Asp Lys Pro
               125                 130                 135
Pro Thr Leu Thr Lys Phe Cys Glu Ser Met Trp Arg Phe Thr Phe
               140                 145                 150
Tyr Leu Cys Ile Phe Cys Tyr Gly Ile Arg Phe Leu Trp Ser Ser
```